

General Video Game AI: a Multi-Track Framework for Evaluating Agents, Games and Content Generation Algorithms

Diego Perez-Liebana, Jialin Liu, Ahmed Khalifa, Raluca D. Gaina, Julian Togelius, Simon M. Lucas

Abstract—General Video Game Playing (GVGP) aims at designing an agent that is capable of playing multiple video games with no human intervention. In 2014, The General Video Game AI (GVGAI) competition framework was created and released with the purpose of providing researchers a common open-source and easy to use platform for testing their AI methods with potentially infinity of games created using Video Game Description Language (VGDL). The framework has been expanded into several tracks during the last few years to meet the demand of different research directions. The agents are required to either play multiples unknown games with or without access to game simulations, or to design new game levels or rules. This survey paper presents the VGDL, the GVGAI framework, existing tracks, and reviews the wide use of GVGAI framework in research, education and competitions five years after its birth. A future plan of framework improvements is also described.

I. INTRODUCTION

Game-based benchmarks and competitions have been used for testing artificial intelligence capabilities since the inception of the research field. For the first four or five decades, such testing was almost exclusively based on board games. However, since the early 2000s a number of competitions and benchmarks based on video games have sprung up. The more well-known include the Ms Pac-Man competition [1], the Mario AI competition [2], the Simulated Car Racing competition [3], the Arcade Learning Environment [4] and the StarCraft [5] competitions.

So far, most competitions and game benchmarks challenge the agents to play a single game (an exception is the Arcade Learning Environment which contains several dozens of games, but so far almost all published studies deal with learning agents for one game at a time). This leads to an over-specialization, or overfitting, of agents to individual games. This is reflected in the outcome of individual competitions – for example, over the more than five years the Simulated Car Racing Competition ran, submitted car controllers got better at completing races fast, but incorporated more and more game-specific engineering and arguably less of general AI and machine learning algorithms. Similarly, the well-performing bots on the StarCraft competitions are generally highly domain – and even strategy-specific, and display very little in the way of learning and adaptation.

It should not come as a surprise that, whenever possible, researchers will incorporate domain knowledge about the game they design an agent for. Yet, this trend threatens to negate the usefulness of game-based AI competitions for spurring and testing the development of stronger and more general AI.

The General Video Game AI competition [6] was founded on the belief that the best way to stop AI researchers from relying on game-specific engineering in their agents is to make it impossible. The idea is that researchers develop their agents without knowing what games they will be playing, and after submitting their agents to the competition all agents are evaluated using an unseen set of games. Every competition event requires the design of a new set of games, as reusing previous games would make it possible to adapt agents to existing games.

In order to make this possible, the Video Game Description Language (VGDL) was developed to easily create new games to be played by agents. VGDL was designed so that it would be easy to create games both for humans and algorithms, eventually allowing for automated generation of testbed games. A game engine was also created to allow games specified in this language to be visualized and played, both by humans and AI agents, forming the basis of the competition.

While the GVGAI competition was initially focused on benchmarking AI algorithms for playing the game, the competition and its associated software has multiple uses. In addition to the competition tracks dedicated to game-playing agents, there are now competition tracks focused on generating game levels or rules. There is also the potential to use VGDL and GVGAI as a game prototyping system, and there is a rapidly growing body of research using this framework for everything from building mixed-initiative design tools to demonstrating new concepts in game design.

This paper gives an overview of the various tracks of the GVGAI competition, the GVGAI software, and the various uses the software has been put to. We start with describing the basics of VGDL and the GVGAI framework. We then describe the various competition tracks that have launched so far: planning (single-player and two-player), level generation, rule generation and learning. The next few sections are devoted to discussing the various kinds of AI methods that have been used in the submissions to each track. Especial consideration is given to the single-player planning track, as it has existed for longest and received the most submissions up to date. This is followed by a section cataloging some of the non-competition research uses of the GVGAI software. The final few sections look forward to the future use and development of the framework and competition: how it can be used in teaching, open research problems (specifically related to the planning tracks), and the future evolution of the competition and framework itself.

II. VGDL AND THE GVGAI FRAMEWORK

The idea of Video Game Description Language (VGDL) was initially proposed by Ebner *et al.* [7] at the Dagstuhl Seminar on Artificial and Computational Intelligence in Games, during which the first VGDL game, Space Invaders, was created. Then, Tom Schaul continued this work by completing and improving the language in a Python framework for model-based learning and released the first game engine in 2013 [8], [9]. VGDL is a text description language that allows for the definition of two-dimensional, arcade, single-player, grid-based physics and (generally) stochastic games and levels. With an ontology defined in the framework, VGDL permits the description of objects (sprites), their properties and the consequences of them colliding with each other. Additionally, termination conditions and reward signals (in form of game scores) can be defined for these games. In total, four different sets are defined: Sprite Set (which defines what kind of sprites take part in the game), Interaction Set (rules that govern the effects of two sprites colliding with each other), Termination Set (which defines how does the game end) and Mapping Set (that specifies which characters in the level file map to which sprites from the Sprite set).

M. Ebner *et al.* [7] and J. Levine *et al.* [10] described, in their Dagstuhl 2013 follow-up, the need and interest for such a framework that could accommodate a competition for researchers to tackle the challenge of General Video Game Playing (GVGP). D. Perez *et al.* [6] implemented a version of T. Schaul’s initial framework in Java and organized the first General Video Game AI (GVGAI) competition in 2014 [11]. In the following years, this framework was extended to accommodate two-player games [12], [13], level [14], rule [15] generation, and real-world physics games [16].

This framework provided an API for creating bots that would be able to play in any game defined in VGDL, without giving them access to the rules of the games nor to the behaviour of other entities defined in the game. However, the agents do have access to a Forward Model (in order to roll the current state forward given an action) during the thinking time, set to 40ms in the competition settings. Controllers also had access at every game tick to game state variables, such game status - winner, time step and score -, state of the player (also referred to in this paper as *avatar*) - position, orientation, resources, health points -, history of collisions and positions of the different sprites in the game identified with a unique *type id*. Additionally, sprites are grouped in *categories* attending to their general behaviour: Non-Player Characters (NPC), static, movable, portals (which spawn other sprites in the game, or behave as entry or exit point in the levels) and resources (that can be collected by the player).

All this information is also provided to agents for the *learning* setting of the framework and competition, with the exception of the Forward Model. In its last (to date) modification of the framework, GVGAI included compatibility for creating learning agents, both in Java and in Python, which would learn to play any game is given in an episodic manner [17]. At the moment of writing, the framework counts on 120 single-player and 60 two-player games.

Contest Leg	Winner	Type	Section
CIG-14	OLETS	Tree Search Method	IV-B [11]
GECCO-15	YOLOBOT	Hyper-heuristic	IV-E [18]
CIG-15	Return42	Hyper-heuristic	IV-E [19]
CEEC-15	YBCriber	Hybrid	IV-D [20]
GECCO-16	YOLOBOT	Hyper-heuristic	IV-E [18]
CIG-16	MaastCTS2	Tree Search Method	IV-B [21]
WCCI-16 (2P)	ToVo2	Hybrid	V-A [13]
CIG-16 (2P)	Number27	Hybrid	V-B [13]
GECCO-17	YOLOBOT	Hyper-heuristic	IV-E [18]
CEC-17 (2P)	ToVo2	Hybrid	V-A [13]

TABLE I

WINNERS OF ALL EDITIONS OF THE GVGAI PLANNING COMPETITION. 2P INDICATES 2-PLAYER TRACK. HYBRID DENOTES 2 OR MORE TECHNIQUES COMBINED IN A SINGLE ALGORITHM. HYPER-HEURISTIC HAS A HIGH LEVEL DECISION MAKER TO DECIDES WHICH SUB-AGENT MUST PLAY (SEE SECTION IV). TABLE EXTENDED FROM [19].

III. GVGAI COMPETITION TRACKS

This section summarizes the different tracks featured in the GVGAI competition.

A. Single-player planning

The first competition track is Single-player Planning [11], in which one agent (also referred to as *bot* or *controller*) plays single-player games with the aid of the Forward Model. Each controller has 1 second for initialization and 40ms at each game tick as decision time.

All GVGAI tracks follow a similar structure: there are several *public* sets of games, included in the framework, allowing the participants to train their agents on them. For each edition, there is one *validation* and one *test* set. Both sets are private and stored in the competition server¹. There are 10 games on each set, with 5 different levels each.

Participants can submit their entries any time before the submission deadline to all training and validation sets, and preliminary rankings are displayed in the competition website (the names of the validation set games are anonymized). All controllers are run on the test set after the submission deadline to determine the final rankings of the competition, executing each agent 5 times on each level.

Rankings are computed by first sorting all entries per game according to victory rates, scores and time steps, in this order. These per-game rankings award points to the first 10 entries, from first to tenth position: 25, 18, 15, 12, 10, 8, 6, 4, 2 and 1. The winner of the competition is the submission that sums more points across all games in the test set. For a more detailed description of the competition and its rules, the reader is referred to [11].

Table I shows the winners of all editions up to date, along with the section of this survey in which the method is included and the paper that describes the approach more in depth.

B. Two-player planning

The Two-player Planning track [12] was added in 2016, with the aim of testing general AI agents in environments

¹www.gvgai.net; Intel Core i5 machine, 2.90GHz, and 4GB of memory.

which are more complex and present more direct player interaction. Games in this setting are played by two-players in a simultaneous move fashion. The rules of the games included in the competition are still secret to the players, similar to the Single-player track, but an additional piece of information is hidden: whether the game is competitive or cooperative. Having both types of games ensures that agents are not tuned to only compete against their opponents, instead having to correctly judge various situations and identify what the goal of the other intelligent presence in the game is. However, this question of an appropriate opponent model remains open, as all competition entries so far have employed random or very simple techniques.

Two legs of this track were organized for the first time in 2016, at the IEEE World Congress on Computational Intelligence (WCCI) and the IEEE Conference on Computational Intelligence and Games (CIG). Although ToVo2 won the first leg and Number27 the second, the winner (adriencx) and runner-up (MaastCTS2) of the overall championship showed that a general agent may not be the best at a particular subset of problems, but will perform at a high level on many different ones. In 2017, ToVo2 did win the overall competition, organized at the IEEE Congress on Evolutionary Computation (CEC). The final results of the competition can be seen in Table I. For more details, the reader is referred to [13].

C. Learning track

The GVGAI Single-Player learning track started in 2017. Among the GVGAI tracks, this is the only one which accepts submission of agent written not only in Java, but also in Python, in order to accommodate for popular machine learning libraries written in this language. The Single-Player Learning track is briefly introduced below. More technical details of the framework and competition procedure are explained in the track technical manual [17].

The learning track is based on the GVGAI framework, but a key difference is that no forward model is provided to the controllers. Hence, learning needs to be achieved by an episodic repetition of games played. Note that, even while no forward model is accessible during the learning and validation phases, controllers receive an observation of the current game state via a *StateObservation* object, which is provided as a Gson object or as a screen-shot of the game screen (in *png* format). The agent is free to select either or both forms of game state observation at any game tick. Similar to the planning tracks, controllers (either in Java or in Python) inherit from an abstract class and a constructor and three methods can be implemented: *INIT* (called at the beginning of every game, must finish in no more than 1s of CPU time), *ACT* (called at every game tick, determines the next action of the controller in 40ms of CPU time) and *RESULT* (called at the end of every game, with no time limit to allow for processing the outcome and perform some learning).

As in the planning track, games sets count on 10 games with 5 levels each. The execution in a given set is split in two phases: a learning phase, which allows training in the first 3 levels of each game, and a *validation* phase, which uses the

Agent	Training set		Test set	
	Score	Ranking	Score	Ranking
kkunan	125	6	184	1
sampleRandom†	154	2	178	2
DontUnderestimateUchiha	149	3	158	3
sampleLearner†	149	4	152	4
ercumentilhan	179	1	134	5
YOLOBOT	132	5	112	6

TABLE II

TABLE SHOWS THE SCORE AND RANKING OF THE SUBMITTED LEARNING AGENTS HAVE OBTAINED ON THE TRAINING AND TEST SET. †DENOTES A SAMPLE CONTROLLER.

Game	1-P Planning	1-P Learning	
	Best score	Best score	Agent
G2	109.00 ± 38.19	31.5 ± 14.65	sampleRandom†
G3	1.00 ± 0.00	0±0	*
G4	1.00 ± 0.00	0.2 ± 0.09	kkunan
G5	216.00 ± 24.00	1±0	*
G6	5.60 ± 0.78	3.45 ± 0.44	DontUnderestimateUchiha
G7	31696.10 ± 6975.78	29371.95±2296.91	kkunan
G8	1116.90 ± 660.84	35.15±8.48	kkunan
G9	1.00 ± 0.00	0.05± 0.05	sampleRandom†
G10	56.70 ± 25.23	2.75 ± 2.04	sampleLearner†

TABLE III

TABLE COMPARES THE BEST SCORES BY SINGLE-PLAYER PLANNING AND LEARNING AGENTS ON THE SAME TEST SET. NOTE THAT ONE OF THE GAMES IN THE TEST SET IS REMOVED FROM THE FINAL RANKING DUE TO BUGS IN THE GAME ITSELF. †DENOTES A SAMPLE CONTROLLER.

other 2 available levels. The execution of a controller in a set has therefore 2 phases:

a) *Learning phase*: In each of the games, each controller has a limited amount of time, 5 minutes, for learning the first 3 levels. It will play each of the 3 levels once then be free to choose the next level to play if there is time remaining. The controller will be able to send an action called *ABORT* to finish the game at any moment. The method *RESULT* is called at the end of every game regardless if the game has finished normally or has been aborted by the controller. Thus, the controller can play as many games as desired, potentially choosing the levels to play in, as long as it respects the 5 minutes time limit.

b) *Validation phase*: Immediately after learning phase, the controller plays 10 times the levels 4 and 5 sequentially. There is no more total time limit, but the agent respects the time limits for *init*, *act* and *result*, and can continue learning during the game playing.

The first GVGAI single-player learning track was held in the IEEE CIG 2017. Besides two sample random agents written in Java and Python and one sample agent using Sarsa [22] written in Java, there were three submissions written in Java and one in Python [23]. Table II illustrates the score and ranking of learning agents on the training and test sets. Table III compares the score achieved by the best planning agent in 2017 on each of the games in the test set and the score achieved by the best learning agent.

D. Level Generation

The Level Generation track [14] started in 2016. The aim of the track is to allow competitors to develop a general level generation algorithm. To participate in this track, competitors must implement their own level generator. Competitors have to

provide at least one function that is responsible for generating the level. The framework provides the generator with all the information needed about the game such as game sprites, interaction set, termination conditions and level mapping. Additionally, the framework also provides access to the Forward Model, in order to allow testing the generated levels via agent simulation. The levels are generated in the form of 2d matrix of characters, with each character representing the game sprites at the specific location determined by the matrix. Competitors have the choice to either use the provided level mapping or provide their own one.

The first level generation competition was held at the International Joint Conference on Artificial Intelligence (IJCAI) in 2016, which received four participants. Each participant was provided a month to submit a new level generator. Three different level generators were provided in order to help the users get started with the system (see Section VII for a description of these). Three out of the four participants were simulation-based level generators while the remaining was based on cellular automata. During the competition day, people attending the conference were encouraged to try pairs of generated levels and select which level they liked (one, both, or neither). Finally, the winner was selected based on the generator with more votes. The winner of the contest was the Easablade generator, a cellular automata described in Section VII-A4.

The competition was run again during IEEE CIG 2017 with same configuration from previous year (one month for implementation followed by on site judging). Unfortunately, only one submission was received, hence the the competition was canceled. This submission used a n-gram model to generate new constrained levels using a recorded player keystrokes.

E. Rule Generation

The Rule Generation track [15] was introduced and held during CIG 2017. The aim of the track is to generate the interaction set and termination conditions for a certain level with a fixed set of game sprites in a fixed amount of time. To participate in the track, competitors have to provide their own rule generator. The framework provides the competitors with game sprites, a certain level, and a forward model to simulate generated games, as in the Level Generation case. The generated games are represented as two arrays of strings. The first array contains the interaction set, while the second array contains the termination conditions.

The first rule generation competition was intended to be run at CIG 2017. Three different sample rule generators were provided (as discussed in Section VIII) and the contest ran over a month's period. Unfortunately, no submissions were received for this track.

IV. METHODS FOR SINGLE PLAYER PLANNING

This section describes the different methods that have been implemented for Single Player Planning in GVGAI. All the controllers that face this challenge have in common the possibility of using the forward model to sample future states from the current game state, plus the fact that they have a limited

action-decision time. While most attempts abide by the 40ms decision time imposed by the competition, other efforts in the literature compel their agents to obey a maximum number of uses of the forward model.

Section IV-A briefly introduces the most basic methods that can be found within the framework. Then Section IV-B describes the different tree search methods that have been implemented for this settings by the community, followed by Evolutionary Methods in Section IV-C. Often, more than one method is combined into the algorithm, which gives place to Hybrid methods (Section IV-D) or Hyper-heuristic algorithms (Section IV-E). Further discussion on these methods and their common take-aways has been included in Section XI.

A. Basic Methods

The GVGAI framework contains several agents aimed at demonstrating how a controller can be created for the single-player planning track of the competition [11]. Therefore, these methods are not particularly strong.

The simplest of all methods is, without much doubt, *doNothing*. This agent returns the action `NIL` at every game tick without exception. The next agent in complexity is *sampleRandom*, which returns a random action at each game tick. Finally, *onestepllookahead* is another sample controller that rolls the model forward for each one of the available actions in order to select the one with the highest action value, determined by a function that tries to maximize score while minimizing distances to NPCs and portals.

B. Tree Search Methods

One of the strongest and influential sample controllers is *sampleMCTS*, which implements the Monte Carlo Tree Search [24] algorithm for real-time games. Initially implemented in a *closed loop* version (the states visited are stored in the tree node, without requiring the use of the Forward Model during the tree policy phase of MCTS), it achieved the third position (out of 18 participants) in the first edition of the competition.

The winner of that edition, Adrien Couëtoux, implemented Open Loop Expectimax Tree Search (OLETS), which is an *open loop* (states visited are never stored in the associated tree node) version of MCTS which does not include rollouts and uses Open Loop Expectimax (OLE) for the tree policy. OLE substitutes the empirical average reward by r_M , a weighted sum of the empirical average of rewards and the maximum of its children r_M values [11].

Torsten Schuster, in his MSc thesis [25], analyzes several enhancements and variations for MCTS in different sets of the GVGAI framework. These modifications included different tree selection, expansion and play-out policies. Results show that combinations of Move-Average Sampling Technique (MAST) and N-Gram Selection Technique (NST) with Progressive History provided an overall higher rate of victories than their counterparts without these enhancements, although this result was not consistent across all games (with some simpler algorithms achieving similar results).

In a different study, Dennis Soemers [21], [26] explored multiple enhancements for MCTS: Progressive History (PH) and NST for the tree selection and play-out steps, tree re-use (by starting at each game tick with the subtree grown in the previous frame that corresponds to the action taken, rather than a new root node), bread-first tree initialization (direct successors of the root node are explored before MCTS starts), safety pre-pruning (prune those nodes with high number of game losses found), loss avoidance (MCTS ignores game loss states when found for the first time by choosing a better alternative), novelty-based pruning (in which states with features rarely seen are less likely to be pruned), knowledge based evaluation [27] and deterministic game detection. The authors experimented with all these enhancements in 60 games of the framework, showing that most of them improved the performance of MCTS significantly and their all-in-one combination increased the average win rate of the sample agent in 17 percentage points. The best configuration was winner of one of the editions of the 2016 competitions (see Table I).

F. Frydenberg studied yet another set of enhancements for MCTS. The authors showed that using MixMax backups (weighing average and maximum rewards on each node) improved the performance in only some games, but its combination with reversal penalty (to penalize visiting the same location twice in a play-out) offers better results than vanilla MCTS. Other enhancements, such as macro-actions (by repeating an action several times in a sequence) and partial expansion (a child node is considered expanded only if its children have also been expanded) did not improve the results obtained.

Perez et al. [27] implemented KB-MCTS, a version of MCTS with two main enhancements. First, distance to different sprites were considered features for a linear combination, where the weights were evolved to bias the MCTS rollouts. Secondly, a Knowledge Base (KB) is kept about how *interesting* for the player the different sprites are, where *interesting* is a measure of curiosity (rollouts are biased towards unknown sprites) and experience (a positive/negative bias for getting closer/farther to beneficial/harmful entities). The results of applying this algorithm to the first set of games of the framework showed that the combination of these two components gave a boost in performance in most games of the first training set.

The work in [27] has been extended by other researchers in the field, which also put a special effort on biasing the Monte Carlo (MC) simulations. In [28], the authors modified the random action selection in MCTS rollouts by using potential fields, which bias the rollouts by making the agent move in a direction akin to the field. The authors showed that KB-MCTS provides a better performance if this potential field is used instead of the Euclidean distance between sprites implemented in [27]. Additionally, in a similar study [29], the authors substituted the Euclidean distance for a measure calculated by a pathfinding algorithm. This addition achieved some improvements over the original KB-MCTS, although the authors noted in their study that using pathfinding does not provide a competitive advantage in all games.

Another work by H. Park and K. Kim [30] tackles this

challenge by a) determining the goodness of the other sprites in the game; b) computing an Influence Map (IM) based on this; and c) using the IM to bias the simulations, in this occasion by adding a third term to the Upper Confidence Bound (UCB) [31] equation for the tree policy of MCTS. Although not compared with KB-MCTS, the resultant algorithm improves the performance of the sample controllers in several games of the framework, albeit performing worse than these in some of the games used in the study.

Biasing rollouts is also attempted by E. dos Santos et al. [32], who introduced Redundant Action Avoidance (RAA) and Non-Defeat Policy (NDP); RAA analyzes changes in the state to avoid selecting sequences of actions that do not produce any alteration on position, orientation, properties or new sprites in the avatar. NDP makes the recommendation policy ignore all children of the root node who found at least one game loss in a simulation from that state. If all children are marked with a defeat, normal (higher number of visits) recommendation is followed. Again, both modifications are able to improve the performance of MCTS in some of the games, but not in all.

M. de Waard et al. [33] introduced the concept of options or macro-actions in GVGAI. Each option is associated with a goal, a policy and a termination condition. The selection and expansion steps in MCTS are modified so the search tree branches only if an option is finished, allowing for a deeper search in the same amount of time. Their results show that Option MCTS outperforms MCTS in games with small levels or a few number of sprites, but loses in the comparison to MCTS when the games are bigger due to these options becoming too large.

In a similar line, Perez et al. [16] employed macro-actions for GVGAI games that used continuous (rather than grid-based) physics. These games have a larger state space, which in turn delays the effects of the player's actions and modifies the way agents navigate through the level. Macro-actions are defined as a sequence or repetition of the same action during M steps, which is arguably the simplest kind of macro-actions that can be devised. MCTS performed better *without* macro-actions on average across games, but there are particular games where MCTS needs macro-actions to avoid losing at every attempt. The authors also concluded that the length M of the macro-actions impacts different games distinctly, although shorter ones seem to provide better results than larger ones, probably due to a more fine control in the movement of the agents.

Some studies have brought multi-objective optimization to this challenge. For instance, Perez et al. [34] implemented a Multi-Objective version of MCTS, concretely maximizing score and level exploration simultaneously. In the games tested, the rate of victories grew from 32.24% (normal MCTS) to 42.38% in the multi-objective version, showing great promise for this approach. In a different study, A. Khalifa et al. [35] apply multi-objective concepts to evolving parameters for a tree selection confidence bounds equation. A previous work by I. Bravi [36] (also discussed later in Section IV-D) provided multiple UCB equations for different games. The work in [35] evolved, using S-Metric Selection Evolutionary

Multi-objective Optimization Algorithm (SMS-EMOA), the linear weights of an UCB equation that results of combining all from [36] in a single one. All these components respond to different and conflicting objectives, and their results show that it is possible to find good solutions for the games tested.

A significant exception to MCTS with regards to tree search methods for GVGAI is that of T. and H. Geffner [20] (winner of one of the editions of the 2015 competition, *YBCriber*, as indicated in Table I), who implemented Iterated Width (IW; concretely IW(1)). IW(1) is a breadth-first search with a crucial alteration: a new state found during search is pruned if it does not make true a new tuple of at most 1 atom, where *atoms* are boolean variables that refer to position (and orientations in the case of avatars) changes of certain sprites at specific locations. The authors found that IW(1) performed better than MCTS in many games, with the exception of puzzles, where IW(2) (pruning according to pairs of atoms) showed better performance. This agent was declared winner in the CEEC 2015 edition of the Single-player planning track [6].

A. Badabi [37] implemented several versions of Enforced Hill Climbing (EHC), a breadth-first search method that looks for a successor of the current state with a better heuristic value. EHC obtained similar results to KB-MCTS in the first set of games of the framework, with a few disparities in specific games of the set.

Finally, Mark J. Nelson [38] ran a study on MCTS in order to investigate if, giving a higher time budget to the algorithm (i.e. increasing the number of iterations), MCTS was able to master most of the games. In other words, if the real-time nature of the GVGAI framework and competition is the reason why different approaches fail to achieve a high victory rate. This study provided up to 30 times more budget to the agent, but the performance of MCTS only increased marginally even at that level. In fact, this improvement was achieved by means of losing less often rather than by winning more games. This paper concludes that the real-time aspect is not the only factor in the challenge, but also the diversity in the games. In other words, increasing the computational budget is not the answer to the problem GVGAI poses, at least for MCTS.

C. Evolutionary Methods

The second big group of algorithms used for single-player planning is that of evolutionary algorithms (EA). Concretely, the use of EAs for this real-time problem is mostly implemented in the form of Rolling Horizon EAs (RHEA). This family of algorithms evolves sequences of actions with the use of the forward model. Each sequence is an individual of an EA which fitness is the value of the state found at the end of the sequence. Once the time budget is up, the first action of the sequence with the highest fitness is chosen to be applied in that time step.

The GVGAI competition includes *SampleRHEA* as a sample controller. *SampleRHEA* has a population size of 10, individual length of 10 and implements uniform crossover and mutation, where one action in the sequence is changed for another one (position and new action chosen uniformly at random) [11].

R. Gaina et al. [39] analyzed the effects of the RHEA parameters on the performance of the algorithm in 20 games,

chosen among the existent ones in order to have a representative set of all games in the framework. The parameters analyzed were population size and individual length, and results showed that higher values for both parameters provided higher victory rates. This study motivated the inclusion of Random Search (*SampleRS*) as a sample in the framework, which is equivalent to RHEA but with an infinite population size (i.e. only one generation is evaluated until budget is consumed) and achieves better results than RHEA in some games. [39] also compared RHEA with MCTS, showing better performance for an individual length of 10 and high population sizes.

A different Evolutionary Computation agent was proposed by B. Jia et al. [40], [41], which consists of a Genetic Programming (GP) approach. The authors extract features from a screen capture of the game, such as avatar location and the positions and distances to the nearest object of each type. These features are inputs to a GP system that, using arithmetic operands as nodes, determines the action to execute as a result of three trees (horizontal, vertical and action use). The authors report that all the different variations of the inputs provided to the GP algorithm give similar results to those of MCTS, on the three games tested in their study.

D. Hybrids

The previous studies feature techniques in which one technique is predominant in the agent created, albeit they may include enhancements which can place them in the boundary of hybrids. This section describes those approaches that, in the opinion of the authors, would in their own right be considered as techniques that mix more than one approach in the same, single algorithm.

An example of one of these approaches is presented by R. Gaina et al. [42] analyzed the effects of seeding the initial population of the RHEA using different methods. Part of the decision time budget is dedicated to initialize a population with sequences that are promising, as determined by *onesteplookahead* and *MCTS*. Results show that both seeding options provide a boost in victory rate when population size and individual length are small, but the benefits vanish when these parameters are large.

Other enhancements for RHEA proposed in [43] are incorporating a bandit-based mutation, a statistical tree, a shift buffer and rollouts at the end of the sequences. The bandit-based mutation breaks the uniformity of the random mutations in order to choose new values according to suggestions given by a uni-variate armed bandit. However, the authors reported that no improvement on performance was noticed. A statistical tree, previously introduced in [44], keeps a game tree with visit count and accumulated rewards in the root node, which are subsequently used for recommending the action to take in that time step. This enhancement produced better results with shorter individuals and smaller population sizes. The shift buffer enhancement provided the best improvement in performance, which consist of shifting the sequences of the individuals of the population one action to the left, removing the action from the previous time step. This variation, similar

to keeping the tree between frames in MCTS, combined with the addition of rollouts at the end of the sequences provided an improvement victory rate (20 percentile points over vanilla RHEA) and scores.

A similar (and previous) study was conducted by H. Horn et al. [45]. In particular, this study features RHEA with rollouts (as in [43]), RHEA with MCTS for alternative actions (where MCTS can determine any action with the exception of the one recommended by RHEA), RHEA with rollouts and sequence planning (same approach as the shift buffer in [43]), RHEA with rollouts and occlusion detection (which removes not needed actions in a sequence that reaches a reward) and RHEA with rollouts and NPC attitude check (which rewards sequences in terms of proximity to sprites that provide a positive or negative reward). Results show that RHEA with rollouts improved performance in many games, although all the other variants and additions performed worse than the sample agents. It is interesting to see that in this case the shift buffer did not provide an improvement in the victory rate, although this may be due to the use of different games.

Torsten Schuster [25] proposes two methods that combine MCTS with evolution. One of them, $(1+1)-EA$ as proposed by [27], evolves a vector of weights for a set of game features in order to bias the rollouts towards more interesting parts of the search space. Each rollout becomes an evaluation for an individual (weight vector), using the value of the final state as fitness. The second algorithm is based on strongly-typed GP (STGP) and uses game features to evolve state evaluation functions that are embedded within MCTS. These two approaches join MAST and NST (see Section IV-B) in a larger comparison, and the study concludes that different algorithms outperform others in distinct games, without an overall winner in terms of superior victory rate, although superior to vanilla MCTS in most cases.

The idea of evolving weight vectors for game features during the MCTS rollouts introduced in [27] (KB-MCTS²) was explored further by Jim van Eeden in his MSc thesis [46]. In particular, the author added A* as a pathfinding algorithm to replace the euclidean distance used in KB-MCTS for a more accurate measure and changing the evolutionary approach. While KB-MCTS used a weight for each pair feature-action, being the action chosen at each step by the Softmax equation, this work combines all move actions on a single weight and picks the action using Gibbs sampling. The author concludes that the improvements achieved by these modifications are marginal, and likely due to the inclusion of pathfinding.

Additional improvements on KB-MCTS are proposed by C. Chu et al. [47]. The authors replace the Euclidean distance feature to sprites with a grid view of the agent’s surroundings, and also the $(1+1)-EA$ with a Q-Learning approach to bias the MCTS rollouts, making the algorithm update the weights at each step in the rollout. The proposed modifications improved the victory rate in several sets of games of the framework and also achieved the highest average victory rate among the algorithms it was compared with.

E. İlhan and A. Etaner-Uyar [48] implemented a combination of MCTS and *true online* Sarsa (λ) [49]. The authors use MCTS rollouts as episodes of past experience, executing true online Sarsa with each iteration with a ϵ -greedy selection policy. Weights are learnt for features taken as the smallest euclidean distance to sprites of each type. Results showed that the proposed approach improved the performance on vanilla MCTS in the majority of the 10 games used in the study.

Evolution and MCTS have also been combined in different ways. In one of them, I. Bravi et al. [50] used a GP system to evolve different tree policies for MCTS. Concretely, the authors evolve a different policy for each one of the (five) games employed in the study, aiming to exploit the characteristics of each game in particular. The results showed that the tree policy plays a very important role on the performance of the MCTS agent, although in most cases the performance is poor - none of the evolved heuristics performed better than the default UCB in MCTS.

Finally, C. Sironi et al. [51] designed three Self-Adaptive MCTS (SA-MCTS) that tuned the parameters of MCTS (playout depth and exploration factor) on-line, using Naive Monte-Carlo, an Evolutionary Algorithm (λ, μ) and the N-Tuple Bandit Evolutionary Algorithm [52]. Results show that all tuning algorithms improve the performance of MCTS there where vanilla MCTS performs poorly, while keeping a similar rate of victories in those where MCTS performs well.

E. Hyper-heuristics / algorithm selection

Several authors have also proposed agents that use several algorithms, but rather than combining them into a single one, there is a higher level decision process that determines which one of them should be used at each time.

Blaine Ross, in his MSc thesis [53] proposes an agent that is a combination of two methods. This approach uses A* with Enforced Hill Climbing to navigate through the game at a high level and switches to MCTS when in close proximity to the goal. The work highlights the problems of computing paths in the short time budget allowed, but indicate that goal targeting with path-finding combined with local maneuvering using MCTS does provide good performance in some of the games tested.

T. Joppen et al. implemented *YOLOBOT* [18], arguably the most successful agent for GVGAI up to date, as it has won several editions of the competition. Their approach consists of a combination of two methods: a heuristic Best First Search (BFS) for deterministic environments and MCTS for stochastic games. Initially, the algorithm employs BFS until the game is deemed stochastic, an optimal solution is found or a certain game tick threshold is reached, extending through several consecutive frames if needed for the search. Unless the optimal sequence of actions is found, the agent will execute an enhanced MCTS consistent of informed priors and rollout policies, backtracking, early cutoffs and pruning. The resultant agent has shown consistently a good level of play in multiple game sets of the framework.

Another hyper-heuristic approach, also winner of one of the 2015 editions of the competition (*Return42*, see Table I),

²This approach could also be considered an hybrid. Given its influence in other tree approaches, it has also been partially described in Section IV-B

determines first if the game is deterministic or stochastic. In case of the former, A* is used to direct the agent to sprites of interest. Otherwise, random walks are employed to navigate through the level [19].

The fact that this type of portfolio agents has shown very promising results has triggered more research into hyper-heuristics and game classification. The work by P. Bontrager et al. [54] used K-means to cluster games and algorithms attending to game features derived from the type of sprites declared in the VGDL description files. The resulting classification seemed to follow a difficulty pattern, with 4 clusters that grouped games that were won by the agents at different rates.

Mendes et al. [55] built a hyper-agent which selected automatically an agent from a portfolio of agents for playing individual game and tested it on the GVGAI framework. This approach employed game-based features to train different classifiers (Support Vector Machines - SVM, Multi-layer Perceptrons, Decision Trees - J48, among others) in order to select which agent should be used for playing each game. Results show that the SVM and J48 hyper-heuristics obtained a higher victory rate than the single agents separately.

H. Horn et al. [45] (described before in Section IV-D) also includes an analysis on game features and difficulty estimation. The authors suggest that the multiple enhancements that are constantly attempted in many algorithms could potentially be switched on and off depending on the game that is being played, with the objective of dynamically adapting to the present circumstances.

D. Ashlock et al. [19] suggest the possibility of creating a classification of games, based on the performance of multiple agents (and their variations: different enhancements, heuristics, objectives) on them. Furthermore, this classification needs to be stable, in order to accommodate the ever-increasing collection of games within the GVGAI framework, but also flexible enough to allow an hyper-heuristic algorithm to choose the version that better adapts to unseen games.

V. METHODS FOR TWO-PLAYER PLANNING

This section approaches agents developed by researchers within the Two-Player Planning setting. Most of these entries have been submitted to the Two-Player Planning track of the competition [12]. Two methods stood out as the base of most entries received so far, Monte Carlo Tree Search (MCTS) and Evolutionary Algorithms (EA) [13]. On the one hand, MCTS performed better in cooperative games, as well as showing the ability to adapt better to asymmetric games, which involved a role switch between matches in the same environment. EAs, on the other hand, excelled in games with long lookaheads, such as puzzle games, which rely on a specific sequence of moves being identified.

Counterparts of the basic methods described in Section IV-A are available in the framework for the Two-Player track as well, the only difference being in the One Step Lookahead agent which requires an action to be supplied for the opponent when simulating game states. The opponent model used by the sample agent assumes they will perform a random move (with

the exception of those actions that would cause a loss of the game).

A. Tree Search methods

Most of the competition entries in the 2016 season were based on MCTS (see Section IV-B).

Some entries employed an Open Loop version, which would only store statistics in the nodes of the trees and not game states, therefore needing to simulate through the actions at each iteration for a potentially more accurate evaluation of the possible game states. Due to this being unnecessarily costly in deterministic games, some entries such as *MaasCTS2* and *YOLOBOT* switched to Breadth-First Search in such games after an initial analysis of the game type, a method which has shown ability to finding the optimal solution if the game lasts long enough.

Enhancements brought to MCTS include generating value maps, either regarding physical positions in the level, or higher-level concepts (such as higher values being assigned to states where the agent is closer to objects it hasn't interacted with before; or interesting targets as determined by controller-specific heuristics). The winner of the 2016 WCCI leg, *ToVo2*, also employed dynamic Monte Carlo roll-out length adjustments (increased with the number of iterations to encourage further lookahead if budget allows) and weighted roll-outs (the weights per action generated randomly at the beginning of each roll-out).

All agents use online learning in one way or another (the simplest form being the base Monte Carlo Tree Search backups, used to gather statistics about each action through multiple simulations), but only the overall 2016 Championship winner, *adrienctx*, uses offline learning on the training set supplied to tune the parameters in the Stochastic Gradient Descent function employed, learning rate and mini batch size.

B. Evolutionary methods

Two of the 2016 competition entries used an EA technique as a base as an alternative to MCTS: *Number27* and *CatLinux* [13].

Number27 was the winner of the CIG 2016 leg, the controller placing 4th overall in the 2016 Championship. *Number27* uses a Genetic Algorithm, with one population containing individuals which represent fixed-length action sequences. The main improvement it features on top of the base method is the generation of a value heat-map, used to encourage the agent's exploration towards interesting parts of the level. The heat-map is initialized based on the inverse frequency of each object type (therefore a lower value the higher the object number) and including a range of influence on nearby tiles. The event history is used to evaluate game objects during simulations and to update the value map.

CatLinux was not a top controller on either of the individual legs run in 2016, but placed 5th overall in the Championship. This agent uses a Rolling Horizon Evolutionary Algorithm (RHEA). A shift buffer enhancement is used to boost performance, specifically keeping the population evolved during one game tick in the next, instead of discarding it, each action

sequence is shifted one action to the left (therefore removing the previous game step) and a new random action is added at the end to complete the individual to its fixed length.

No offline learning used by any of the EA agents, although there could be scope for improvement through parameter tuning (offline or online).

C. Opponent model

Most agents submitted to the Two-Player competition use completely random opponent models. Some entries have adopted the method integrated within the sample *One Step Lookahead* controller, choosing a random but non-losing action. In the 2016 competition, *webpigeon* assumed the opponent would always cooperate, therefore play a move beneficial to the agent. *MaasCTS2* used the only advanced model at the time: it remembered Q-values for the opponent actions during simulations and added them to the statistics stored in the MCTS tree nodes; an ϵ -greedy policy was used to select opponent actions based on the Q-values recorded. This provided a boost in performance on the games in the WCCI 2016 leg, but it did not improve the controller’s position in the rankings for the following CIG 2016 leg.

Opponent models were found to be an area to explore further in [13] and Gonzalez and Perez looked at 9 different models integrated within the sample MCTS agent provided with the framework [56]. *Alphabeta* builds a tree incrementally, returning the best possible action in each time tick, while *Minimum* returns the worst possible action. *Average* uses a similar tree structure, but it computes the average reward over all the actions and it returns the action closest to the average. *Fallible* returns the best possible action with a probability $p = 0.8$ and the action with the minimum reward otherwise. *Probabilistic* involved offline learning over 20 games in the GVGAI framework in order to determine the probability of an MCTS agent to select each action, and then using these to determine the opponent action while playing online. *Same Action* returns the same action the agent plays, while *Mirror* returns its opposite. Finally, *LimitedBuffer* records the last $n = 20$ actions performed by the player and builds probabilities of selecting the next action based on this data, while *UnlimitedBuffer* records the entire history of actions during the game. When all 9 opponent models were tested in a round robin tournament against each other, the probabilistic models achieve the highest win rates and two models, *Probabilistic* and *UnlimitedBuffer* outperforming a random opponent model.

VI. METHODS FOR SINGLE-PLAYER LEARNING

The GVGAI framework has also been used from an agent learning perspective. In this setting, the agents do not use the forward model to plan ahead actions to execute in the real game. Instead, the algorithms learn the games by repeatedly playing them multiple times (as *episodes* in Reinforcement Learning), ideally improving their performance progressively.

This section first describes those approaches that tackled the challenge set in the single-player learning track of the competition (described in Section III-C), to then move to those the ones not adhered to the competition format (Section VI-B.

A. 2017 competition entries

Some of the entries are available in the framework³.

1) *Random agent*: A sample random agent, which selects an action at uniform random at every game tick, is included in the framework (in both Java and Python) for the purposes of testing. This agent is also meant to be taken as a baseline: a learner is expected to perform better than an agent which acts randomly and does not undertake any learning.

2) *Multi-armed bandit algorithms*: *DontUnderestimateUchiha* by K. Kuanusont is based on two popular Multi-Armed Bandit (MAB) algorithms, ϵ -Decreasing Greedy Algorithm [57] and UCB [31]. At any game tick T , the current *best* action a^* with probability $1 - \epsilon_T$ is picked, otherwise an action is uniformly randomly selected. The *best* action a^* at time T is defined as in Equation 1.

$$a^* = \arg \max_{a \in A} \left(\Delta \hat{score}_a + \sqrt{\frac{2 \log T}{t_a}} \right), \quad (1)$$

where t_a denotes the number of times that the action a has been selected, and $\Delta \hat{score}_a$ denotes the empirical mean increment of score by applying the action a so far.

This is a very interesting combination, as the UCB-style selection (Equation 1) and the ϵ -Decreasing Greedy Algorithm both aim at balancing the trade-off between exploiting more the best-so-far action and exploring others. Additionally, ϵ_0 is set to 0.5 and it decreases slowly along time, formalized as $\epsilon_T = \epsilon_0 - 0.0001T$. According to the competition setting, all games will last longer than 2,000 game ticks, so $\forall T \in \{1, \dots, 2000\}$, $0.5 \geq \epsilon_T \geq 0.3$. As a result, random decisions are made for approximately 40% time.

3) *State-Action-Reward-State-Action*: *sampleLearner* and *ercumentilhan* are based on the State-Action-Reward-State-Action (SARSA) algorithm. Both agents choose to use a subset of the whole game state information to build a new state to reduce the amount of information to be saved and to take into account similar situations. The main difference is that the former uses a square region with fixed size centered at the avatar’s position, while the latter uses a first-person view with a fixed distance.

4) *Q-learning*: *kkunan*, by K. Kuanusont, is a simple Q-learning agent using most of the avatar’s current information as features, which a few exceptions (as avatar’s health and screen size, as these elements that vary greatly from game to game). The reward at game tick $t + 1$ is defined as the difference between the game score at game tick $t + 1$ and the one at t . The learning rate α and discounted factor γ are manually set to 0.05 and 0.8. During the *learning phase*, an random is performed with probability $\epsilon = 0.1$, otherwise, the best action is selected. During the *validation phase*, the best action is always selected. Despite it’s simplicity, it won the only edition of this track up to date, being the only entry that ranked above the sample random agent.

5) *Tree search methods*: *YOLOBOT* is an adaption of the *YOLOBOT* planning agent (as described previously in Section IV-E). As the forward model is no more accessible

³<https://github.com/GAIGResearch/GVGAI/tree/master/clients/>

in the learning track, the MCTS is substituted by a greedy algorithm to pick the action that minimizes the distance to the chosen object at most. According to the authors, the poor performance of *YOLOBOT* in the learning track, contrary to its success in the planning tracks, was due to the collision model created by themselves that did not work well.

B. Other learning agents

One of the first works that used this framework as a learning environment was carried out by S. Samothrakis et al. [58], who employed Neuro-Evolution in 10 games of the benchmark. Concretely, the authors experimented with Separable Natural Evolution Strategies (S-NES) using two different policies (ϵ -greedy versus softmax) and a linear function approximator versus a neural network as a state evaluation function. Features like score, game status, avatar and other sprites information were used to evolve learners during 1000 episodes. Results show that ϵ -greedy with a linear function approximator was the better combination to learn how to maximize scores on each game.

A. Braylan and R. Miikkulainen [59] performed a study in which the objective was to learn a forward model on 30 games. The objective was to learn the next state from the current one plus an action, where the state is defined as a collection of attribute values of the sprites (spawns, directions, movements, etc.), by means of logistic regression. Additionally, the authors transfer the learnt object models from game to game, under the assumption that many mechanics and behaviours are transferable between them. Experiments showed the effective value of object model transfer in the accuracy of learning forward models, resulting in these agents being stronger at exploration.

Also in a learning setting, K. Kunanusont et al. [60], [61] developed agents that were able to play several games via screen capture. In particular, the authors employed a Deep Q-Network [62] in 7 games of the framework of increasing complexity, and included several enhancements to GVGAI to deal with different screen sizes and a non-visualization game mode. Results showed that the approach allowed the agent to learn how to play in both deterministic and stochastic games, achieving a higher winning rate and game score as the number of episodes increased.

VII. METHODS FOR LEVEL GENERATION

Different researchers used different approaches to generate levels for the GVGAI framework. The following subsection describes all known generators either included in the framework or developed during the competition.

A. Constructive methods

Constructive generators are designed to generate the level based on general knowledge. For example: enemies should be away from the avatar, walls shouldn't divide the world into islands, etc. Based on the game the generator adjusts a couple of parameters and rules to fit the game as, for example, the number of non-playable characters (NPCs) in the generated

level. Constructive generators don't need any simulations after generating the level. The following are the known constructive generators.

1) *Sample random generator*: This is the most naive method to generate a level. The generator first identifies solid sprites. Solid sprites block the avatar and all NPCs from moving and don't react to anything. The generator adds a selected solid sprite as a border for the generated level to prevent sprites from wandering outside the game screen. Followed by adding one of each character in the level mapping section to a random location in the level. This step ensures the game is playable. Finally, it adds a random amount of random sprites from the level mapping to random locations in the level.

2) *Sample constructive generator*: This generator uses some general game knowledge to generate the level. First, the generator calculates the level dimensions and the number of sprites in the level, then labels game sprites based on their interactions and sprite types. After that, it constructs a level layout using the solid sprites, to later add the avatar to a random empty location. After knowing the avatar position, the generator adds harmful sprites (those that can kill the avatar) in a far location from the avatar and adds other sprites at any random free locations. Finally, the generator makes sure that the number of goal sprites is sufficient to prevent winning or losing automatically when the game starts.

3) *Easablade constructive generator*: This is the winner generator for the first level generator competition. The generator is similar to the sample constructive generator but it uses cellular automata to generate the level instead of layering the objects randomly. The cellular automata is run on multiple layers. The first layer is to design the map obstacles, followed by the exit and the avatar, then the goal sprites, harmful sprites, and others.

4) *N-Gram constructive generator*: This generator uses a n-gram model to generate the level. The generator records the player actions from a previous play-through. This action sequence is used to generate the levels using predefined rules and constraints. For example, if the player uses the USE action quite often, the generator will include more enemies in the level. The n-gram is used to specify the rules. Instead of reacting to each separate action, the model reacts to a n-sequence of actions. During the generation process, the algorithm keeps track of the number and position of every generated object to ensure the generated sprites do not overpopulate the level. A single avatar sprite is placed in the lower half of the level.

B. Search-based methods

Search-based generators use simulations to make sure the generated level is playable and better than just placing random objects. The following are the known search-based generators.

1) *Sample genetic generator*: This is a search-based level generator based on the Feasible Infeasible 2 Population Genetic Algorithm (FI2Pop) [63]. FI2Pop is a genetic algorithm which uses 2 populations, one for feasible chromosomes and the other for infeasible chromosomes. The feasible population tries to increase the difference between the OLETS agent (see Section IV-B) and one-step look ahead, while the infeasible

population tries to decrease the number of chromosomes that violate the problem constraints (i.e. at least one avatar in the game, the avatar must not die in the first 40 steps, etc.). Each population evolves on its own, where the children can transfer between the two populations. This generator initializes the population using sample constructive generator.

2) *Amy12 genetic generator*: This generator is built on top of the sample genetic generator. The main idea is to generate a level that fits a certain suspense curve. Suspense is calculated at each point in time, by calculating the number of actions that leads to death or tie using the OLETS agent. The algorithm modifies the levels to make sure the suspense curve is not constant during the life time of the game. Good generators are aimed at producing 3 suspense peaks with values of 50% (where half of the actions, on average lead to losing the game). One of the advantages of using this technique that it makes sure that the generated level is winnable. Games that are hard to win will have a higher peak in the suspense curve, which is not valued highly by the generator.

3) *Inicho genetic generator*: This generator [64] uses a standard genetic algorithm with similar crossover and mutation operators to the sample genetic algorithm. The fitness function used is a combination between the score difference and the constraints specified in the sample genetic generator. The score difference is calculated between an Monte Carlo Tree Search agent and One Step Look Ahead agent. The score difference is normalized between 0 and 1 to make sure it won't overshadow the constraint values.

4) *Number13 genetic generator*: This is a modified version of the sample genetic generator. These modifications includes using adaptive crossover mechanism, adaptive mutation rate, a better agent than OLETS, and allowing crossover between feasible and infeasible population, which is not allowed in the sample genetic generator.

C. Constraint-based methods

1) *ASP generator*: This generator [65] uses Answer Set Programming (ASP) to generate levels. The main idea is to generate ASP rules that generate suitable levels for the current game. The generated rules consists of three different types. The first type are basic rules, which are based on specific decisions to keep the levels simple (for instance, levels can only have one sprite per tile). The second type are game specific rules, which are extracted from the game description file. An example is the identification of singleton sprites that should only have one sprite in the level. The last type are additional rules to minimize the search space. These rules limit the minimum and maximum number of each sprite type. All the rules are evolved using evolutionary strategy with the algorithm performance difference between sampleMCTS and a random agent as the fitness function.

VIII. METHODS FOR RULE GENERATION

This section describes the different algorithms that are included in the framework or have been found in the literature [66] toward generating rules for the GVG-AI framework.

A. Constructive methods

Constructive methods are algorithms that generate the rules in one pass without the need to play the game. The constructive methods often incorporate knowledge about game design to generate more interesting games.

1) *Sample random generator*: This is the simplest generator provided with the framework. The main idea is to generate a game that compiles with no errors. For example, the game shouldn't contain interactions such as killing the end of screen (EOS) sprite. The algorithm starts by generating a random number of interactions by selecting two random sprites (including EOS) and a random interaction rule one by one. The algorithm checks that every interaction is valid before adding it to the generated game. After generating the random interactions, the algorithm generates two termination conditions, one for winning and one for losing. The losing condition is fixed to the avatar being killed, while the winning is either winning the game after a random amount of frames or winning the game when the number of a certain sprite reaches zero.

2) *Sample constructive generator*: This is a more complex generator that utilizes knowledge about VGDL language and level design to generate more interesting games. The algorithm starts by classifying the game sprites into different categories, such as wall sprites (sprites that surround the level), collectible/harmful sprites (immovable sprites that cover around 10% of the level), spawner sprites (sprites that spawn another sprite type), etc. For each type of sprite, the algorithm has rules to generate interactions based on them. For example, harmful sprites kill the avatar on collision, wall sprites either prevent any movable object from passing through or kill the movable object upon collision, etc. For more details about the rules, the reader is referred to [15]). After the game interactions are generated, two termination conditions are generated, one for winning and one for losing. The losing condition is fixed to the avatar's death, while the winning condition depends on the current sprites. For example: if collectible sprites exist in the current definition, the winning condition is set to collect all of them.

B. Search-based methods

Search-based methods use a search based algorithm to find a game based on certain criteria that ensure the generated game have better rules than just randomly choosing them.

1) *Sample genetic generator*: Similar to the level generation track, the search based algorithm uses Feasible Infeasible 2 Population Genetic Algorithm (FI2Pop) [63] to evolve new games. As discussed before, FI2Pop keeps two populations one for feasible games and the other for infeasible games. The infeasible games tries to become feasible by satisfying multiple constraints such as minimizing the number of bad frames (frames contains sprites outside the level boundaries) under certain threshold, the avatar doesn't die in the first 40-frames, etc. On the other hand, the feasible chromosomes try to maximize its fitness. The fitness consists of two parts, the first part is to maximize the difference in performance between the OLETS and MCTS agents, and the difference between

MCTS and random agent. The second part is to maximize the number of interaction rules that fires during the simulation of the generated game.

2) *Thorbjrn generator*: This generator [66] is inspired by the sample genetic generator. Similar to the sample genetic generator, it tries to maximize the difference between the performance of different algorithms. This generator uses evolutionary strategies with mutation and crossover operators to generate an entire game instead of an interaction set and termination conditions.

IX. RESEARCH THAT BUILDS ON GVGAI

A. AI-assisted game design

Machado et al. [67] implemented a recommender system based on the VGDL to recommend game elements, such as sprites and mechanics. Then, the recommender system was expanded to *Cicero* [68], an AI-assisted game design tool built on top of the GVGAI. *Cicero* has a statistics tool of the interactions to help figuring out the unused game rules; a visualization system to illustrate the information about game objects and events, a mechanics recommender, a query system for in-game data and a retrospective analysis application *Seek-Whence* [69]. The gameplay sessions by human players or AI agents can be recorded and every single frame at every game tick can be easily extracted for further study and analysis.

Recently, Liu et al. [70] applied a simple Random Mutation Hill Climber (RMHC) and a Multi-Armed Bandit RMHC together with resampling methods to tune game parameters automatically. Games instances with significant skill-depth have been evolved using GVGAI agents. Furthermore, Kuanusont et al. [52] evolved simultaneously the GVGAI agents as part of the game (opponent models).

Guerrero et al. [71] explored five GVGAI agents using four different heuristics separately on playing twenty GVGAI games, allowing different behaviors according to the diverse scenarios presented in the games. In particular, this work explored heuristics that were not focused on winning the game, but to explore the level or interact with the different sprites of the games. These agents can be used to evaluate generated games, thus help evolve them with preferences to particular behaviors.

Khalifa et al. [72] modified MCTS agents by editing the UCT formula used in the agent. Human playing data has been used for modeling to make the modified agents playing in a human-like way. Primary results showed that one of the studied agents achieved a similar distribution of repeated actions to the one by human players. The work was then extended by Bravi et al. [50], in which game-play data have been used to evolve effective UCT alternatives for a specific game. The MCTS agents using new formulas, with none or limited domain information, are compared to a standard implementation of MCTS (the *sampleMCTS* agent of GVGAI) on the game *Missile Command*. Applying the UCT alternatives evolved using game-playing data to a standard MCTS significantly improved its performance.

Besides designing games and the agents used in them, the automatic generation of video game tutorials (aimed at

helping players understanding how to play a game) is also an interesting sub-field of study. Green et al. [73] pointed out that the GVGAI Framework provides an easy testbed for tutorial generation. The game rules in GVGAI are defined in VGDL, therefore the tutorial generation can be easily achieved by reading and translating VGDL files.

A more recent work by Anderson et al. [74] focused on designing deceptive games to deceive AI agents and lead the agents away from a globally optimal policy. Designing such games helps understand the capabilities and weaknesses of existing AI agents and can serve at a preparation step for designing a meta-agent for GVGP which combines the advantages of different agents. The authors categorized the deceptions and imported various types of deception to the existing GVGAI games by editing the corresponding VGDL files. The agents submitted to the GVGAI single-player playing competition have been tested on the new games. Interestingly, the final ranking of the agents on each of the games differed significantly from the rankings in the GVGAI competition. The new designed deceptive games successfully explored the weaknesses of agents which have performed well on the test set of the official competition.

B. Game generation with RAPP

Nielsen et al. [75] proposed Relative Algorithm Performance Profile (RAPP) as a measure of relative performance of agents and tested their approach on different general game-playing AI agents using GVGAI framework. The authors showed that well-designed games have clear skill-depth, thus being able to distinct good or bad players. In other words, a strong agent or human player should perform significantly better than a weak agent or human player over multiple playings on well-designed games. For instance, a skillful agent is expected to perform better than a random agent, or one that does not move.

Then, Nielsen et al. [66] integrated the differences of average game scores and win rate between any agent and a random agent to the evaluation of new games either randomly generated or generated by editing existing GVGAI games. Though most of the resulted games are interesting to play, there are some exceptions, in which the core challenge of the game has been removed. For instance, the enemy can not hurt the player, which makes it no more an enemy. But it still provides useful starting points for human designers.

Kuanusont et al. [52] extended the idea of RAPP. Five GVGAI agents and a deterministic agent designed for the tested Space Battle Game are used as the candidate opponent, which is considered as part of the game to be evolved. Two GVGAI agents, One Step Look Ahead (weak), MCTS (strong) and the deterministic agent (mediocre), are used to play multiple times the evolved game for evaluation. The evaluation function is defined as the minimum of the difference of game scores between the strong and mediocre agents, and the difference of game scores between the mediocre and weak agents, aiming at generating games that can clearly distinguish stronger agents and weak agents. A recently proposed N-Tuple Bandit Evolutionary Algorithm [76] has been compare to the

simple RMHC. Two human players have tested the generated games and preferred the new game evolved using the N-Tuple Bandit Evolutionary Algorithm.

C. Robustness testing

Perez et al. [77] ran a study on the winners of the 2014 and 2015 editions of the single player planning competition in order to analyze how robust they were to changes in the environment with regards to actions and rewards. The aim of this work was to analyze a different type of generality: controllers for this framework are developed to play in multiple games under certain conditions, but the authors investigated which could be the effect of breaking those compromises. An inaccurate and noisy Forward Model, an agent that does not execute the move decided by the algorithm, or score penalties incurred by performing certain actions.

An interesting conclusion on this study is that, once the conditions have been altered, sample agents climb up to the top of the rankings and the good controllers behave worse. Agents that rely on best first search or A* (such as *YOLOBOT* or *Return42*, already described in this paper) handled noise very badly. MCTS also showed to be quite robust in this regard, above other Rolling Horizon agents that could not cope so well with these changes. This work also reinforced the idea that the GVGAI framework and competition are also robust. Despite the changes in the performance of the agents, some controllers do better than others under practically all conditions. The opposite (rankings depending only on noise factors, for instance) would mean that the framework is quite fragile.

X. EDUCATIONAL USE OF GVGAI

The GVGAI framework has been used to provide engaging assignments for taught modules, and as the basis for many MSc dissertation projects. The descriptions below give an idea of the educational uses of GVGAI but are not intended to be an exhaustive list.

A. Taught Modules

GVGAI has been used in at least two distinct ways within taught modules. The most typical way is to use design specific aspects of the course around the framework, teaching the students about the main concepts of GVGAI with examples of how to write agents for the selected tracks. This is then followed up with an assignment, where a significant weight is given to how well each student or group's entry performs in the league. Several institutions have run private leagues for this, including Otto Von Guericke Universität Magdeburg, University of Essex, University of Muenster, Universidad Carlos III de Madrid, Universidad de Malaga and New York University. Running a private league means the course supervisor has full control over the setup of the league, including when students can enter and how thoroughly the entries are evaluated, and the set of games to evaluate them on. For the 2-player track, this also allows control over the opponents chosen.

Another use-case in taught modules is to teach the VGDL part of framework, and then set the development of novel and

interesting games as the assignment. This was done to good effect at IT University of Copenhagen, where the students produced a number of challenging puzzle games that were later used in the planning track training and validation sets. A similar approach was taken in a module on AI-Assisted Game Design at the University of Essex, where planning track games were also produced.

B. MSc Dissertation Projects

GVGAI offers an extensive range of interesting research challenges, some of which have been addressed in MSc dissertation projects. The majority of the ones we are aware of have focused on the planning track, but this is not surprising as it was the first track to be developed. The planning track also has the benefit of providing some good sample agents as starting points for further work either in the sense of extending the samples to achieve higher performance, or using the samples as a useful source of comparison. A good example is the work on MCTS with options, which used showed the the version with options significantly outperformed the sample MCTS agent on most of the games studied: as with many cases what began as an MSc thesis was later published as a conference paper [33]. In our experience this usually provides an excellent educational experience for the student. Other planning track theses include [25], the real-time enhancements of [26], knowledge-based variants [46] and goal-oriented approaches [53].

Beyond the planning track, other examples (already described in this survey) include applying Answer-Set Programming (ASP) [78] or Genetic Algorithms (GA) [64] for the level generation track and learning from screen capture [60] (this was essentially a learning track approach before the learning track was running). Finally, another approach is to extend the framework in some way, such developing the 2-player learning track [79].

XI. DISCUSSION AND OPEN RESEARCH PROBLEMS ON SINGLE AND TWO-PLAYER PLANNING

The single and two-player planning versions of GVGAI are the ones that have received most attention and research. Despite their popularity and efforts, the best approaches rarely surpass an approximately 50% victory rate in competition game sets, with very low victory percentage in a great number of games. Similarly, different MCTS and RHEA variants (including many of the enhancements studied in the literature) struggle to achieve a higher than 25% victory rate in all (more than a hundred) single-player games of the framework. Therefore, increasing performance in a great proportion of games is probably the most challenging problem at the moment.

Literature shows multiple enhancements on algorithms and methods aiming to improve this performance, but in the vast majority of cases the improvements only affect a subset of games or certain configurations of the algorithms. While this is understandable due to the nature of general video game playing, it also shows that the current approaches does not work in order to reach truly general approaches that work across board.

The work described in this survey has shown, however, interesting insights that can point us in the right direction. For instance, several studies show that using more sophisticated (i.e. with A* or other methods such as potential fields) distances to sprites as features works better than Euclidean distances. The downside is that computing these measurements take an important part of the decision time budget, which can't be used in case it is needed for some games or states where the best action to take is not straight-forward.

In general, one could say that one of the main points to address is how to use the decision time more wisely. Some approaches tried to make every use of the forward model count, like those agents that attempt to learn facts about the game during the rollouts of MCTS. Again, some attempts in this direction have provided marginal improvements, but the problem may be trying to a general feature *extractor*. In other words, what we try to learn is influenced by what we know about existing games (i.e. some sprites are good, other are bad, some spawn other entities and there are things - resources - that the avatar can collect). Some games may require features that have not been thought of, especially because the challenge itself presents games that has not been seen before to the participants.

Another improvement that has been tried in several studies is the use of macro-actions (in most cases, a repetition of an action during several consecutive steps) to a) make the action space coarser; and b) make a better use of the time budget. Again, these modifications have improved performance in certain games (including some that had not been won by any algorithm previously) but they either did not have an impact in others, or they made the performance worse. It is likely that different games can benefit from different macro-action lengths (so work could be done on trying to automatically and dynamically adapt the number of times the action is repeated) but also of more complex structures that allow for high-level planning. In fact, games that require high-level planning are still and open problem to be solved in this setting.

Games classification and the use of hyper-heuristics are also an interesting area for research. Some of the best approaches up to date, as YOLOBOT, do make a differentiation between stochastic and deterministic games to later use one or another algorithm. An open challenge is how to make this classification more accurate and detailed, so an approach could count on a portfolio of (more than 2) algorithms that adapt to every game. Attempts have been made to classify with game features, but results suggest that these classifications and the algorithms used are not strong enough. Devising more general features for this clustering, maybe focused on the agent gameplay experience rather than game features, is a line of future research.

All these unsolved issues apply to both single and two player settings, although the latter case adds the difficulty of having an opponent to compete or collaborate with. There are two open problems that arise from this: first, no study has been made that tries to identify the game and behaviour of the opponent as collaborative or competitive. Analysis of the other player's intentions can be seen as a sub-field on its own, only that in this case we add the general game playing component

to it. Secondly, some advancements have been done in using opponent models that go beyond random, but investigation in more complicated opponent models that better capture and learn the behaviour of the other player could potentially yield better results.

Beside the development of agents for game playing, AI-assisted game design and automatic game testing and debugging using GVGAI agents have attracted researchers' attention. Some work around evolving game skill-depth using relative performance between GVGAI agents have been done recently, and most of this work has been focused on Relative Algorithm Performance Profiles (RAPP), where *performance* is measured in terms of how well the agents play the given games. However, it is sensible to explore other aspects of agent game play to influence game design. Factors like the amount of level explored by different agents (so a generator favours those levels or games that allow for a wider exploration, or maybe a progressive one), their decisiveness [80] on selecting the best action to take or the entropy of their moves can also be used to this end.

XII. FUTURE DIRECTIONS

The GVGAI framework and competition are in constant development. The opportunities that this benchmark provides for different lines of research and education are varied, and this section outlines the future directions planned ahead for the following years.

A. New tracks

As new challenges are proposed, the possibility of organizing them as tracks for the competition arise. Below are listed some possible new tracks that can attract interesting areas of research.

1) *Automatic game design*: The game design involves, but not limited to, game generation, level generation, rule generation and playtesting (playing experience, game feeling, fun, etc.), study of game market, user interface design and audio design. The automatic game design becomes an active research topic since the late 2000's, though, to the best of our knowledge, Barney Pell firstly automatically generated rules for chess-like games in 1992 [81]. A review of the state of the art in automatic game design can be found in [70].

A *Game Generation track* would aim at providing AI controllers which automatically generate totally new games or game instances by varying the game parameters, i.e., parameter tuning. How to achieve the former is an open question. The straight-forward way would be providing a particular theme, a database of game objects, or searching spaces of game rules, with which the participants can generate new games. The ideal case would be that the controllers automatically create totally new games from nothing. Though there is a yawning gulf between aspiration and reality, an interdisciplinary field combining automatic game design and domain-specific automatic programming is expected. The latter, automatic game tuning, is relatively easier. Some search-based and population-based methods have been applied to game parameter optimisation aiming at maximising the depth of game variants [70] or finding more playable games [82].

2) *Multi-Player GVGAI*: Multi-agent games has drawn people’s attention, for instance, real time strategy games (e.g. StarCraft) and board games (e.g. Mahjong). The study of multi-agent GVGAI is a fruitful research topic. Atari games can also be extended to multi-agent games. In particular, the Pac-Man can be seen as a multi-agent game and related competitions have been held since 2011 [1]. The most recent Ms Pac-Man vs Ghost Team Competition [83], which included partial observability, was held at CIG in 2016. Nevertheless, a more general multi-agent track is favorable.

The interface of the Two-Player Planning Track was initially developed for two or more players, so it has the potential to be expanded to a Multi-player planning track, in which an agent is allowed to control more than one player or each of the players is controlled by a separate agent. This future track can be expanded again as a multi-agent learning framework, providing a two-or-more player learning track.

3) *Turing Test GVGAI*: Determining if an agent that is playing a game is a human or a bot is a challenge that has been subject of study for many years [84], [85], and the idea of applying it to a general video game setting is not new [86]. This offers an interesting opportunity to extend the framework to having a Turing Test track where participants create AI agents that play like humans for any game that is given. Albeit the understandable difficulty of this problem, the interest for research in this area is significant: what are the features that can make an agent play like a human in any game?

B. General directions

There are several improvements and additions to the framework that can be done and would potentially affect all existent and future competition tracks. One of these continuous modifications is the constant enlargement of the games library. Not only new games are added for each new edition of the competition, but the work done on automatic game design using the GVGAI framework has the potential to create infinite number of games that can be integrated perfectly into the framework.

Adding more games can also be complemented with compatibility with other systems. Other general frameworks like OpenAI Gym [87], ALE [4] or Microsoft Malmö [88] count on a great number of single or multi-player, model-free or model based tasks. Interfacing with these systems would increment the number of available games which all GVGAI agents could play via a common API. This would also open the framework to 3D games, an important section of the environments the current benchmark does not cover.

With regards to the agents, another possibility could be to provide them with a wider range of available actions. For instance, the player could be able to apply more than one action simultaneously, or these actions could form a continuous action space (i.e. pressing a throttle in a range between 0 and 1). In general, this would enhance the number of legal combinations for the agent to choose from at each decision step.

Beside the framework itself, the website for GVGAI could also be improved to provide better and faster feedback to

the competition participants. More data analysis features can be added, such as visualization of the score changes during the game playing, the action entropy and the exploration of the game world (heat-map of visited positions). A related work is to provide better and more flexible support for game play metric logging, better support for data mining of results together with visualization, and better data saving, which will help enabling to upload replays (i.e., action logs) from AI agents and human play-throughs.

Another envisaged feature is being able to play the game in a web browser (without any download or installation) by an AI agent or human, and visualize the analyzed features during the game playing in real time. A bonus will be the easy parametrisation options for games., thus a player or an AI agent can easily set up the parameters and rules to define the desired game by inserting values directly or generate pseudo-randomly a level to play using some pre-implemented automatic game tuning techniques given some particular goals or features.

XIII. CONCLUSIONS

The GVGAI framework offers the most comprehensive system to date for evaluating the performance of general video game playing agents, and for testing general purpose algorithms for creating new games and creating new content for novel games. The framework has been used in multiple international competitions, and has been used to evaluate the performance of hundreds of general video game agents.

The agent tracks cater for planning agents able to exploit a fast forward model, and learning agents that must learn to react sensibly without the benefits of a forward model. The planning track already comes in single and two-player versions; the learning track is currently single-player only, but with a two-player version under development. Although long-term learning may also be used within the planning track, the best-performing agents have, as far as we know, not yet done this. The success of AlphaGoZero [89] indicates what can be achieved by combining learning and planning, so applying a similar system within GVGAI is an interesting prospect.

The main alternative to GVGAI is the Atari Learning Environment (ALE). At the time of writing, ALE offers higher-quality games than GVGAI as they were home-console commercial games of a few decades ago. In GVGAI terms, ALE offers just two tracks: single-player learning, and single-player planning, with the learning track being the more widely used. For future work on machine learning in video games, we predict that the 2-player tracks will become the most important, as they offer open-ended challenges based on an arms race of intelligence as new players are developed, and are also outside of the current scope of ALE.

In addition to offering a wider range of tasks, GVGAI also benefits from being much more easily extensible than ALE, though ALE has so far had greater uptake within some sectors of the machine learning community. An immediate priority is to test the rich set of ALE agents on the equivalent GVGAI-tracks to gain a sense of the relative difficulty of each environment and to learn more of the relative challenges offered by each one.

The content creation tracks offer an extremely hard challenge: creating rules or levels for unseen games. Promising directions include the further development and exploitation of a range of general game evaluation measures [80], and greater use of the best GVGAI agents to perform the play-testing of the novel rules and levels.

The Video Game Description Language (VGDL) has been an important part of GVGAI to date, since it makes it possible to rapidly and concisely specify new games. However, it is also a source of limitation, as its limited expressivity makes it hard to make games which are fun for humans to play. VGDL also limits the ease with which complex game mechanics can be embedded in games, which in turn limits the depth of challenge that can be posed for the GVGAI agents. Hence an important future direction is the authoring of GVGAI-compatible games in any suitable language which conform to the necessary GVGAI API in order to ensure compatibility with the desired GVGAI track.

Finally, while the above discussion provides a compelling case for the future of GVGAI as a tool for academic study, we also believe that when it reaches a higher level of maturity it will provide an important tool for game designers. The vision is to provide an army of intelligent agents with a range of play-testing abilities, and a diverse set of metrics with which to analyse a range of important functional aspects of a game.

ACKNOWLEDGEMENTS

The authors would like to thank the participants of all tracks of the competition for their work and submitted controllers and generators. This work was partially funded by the EPSRC CDT in Intelligent Games and Game Intelligence (IGGI) EP/L015846/1.

REFERENCES

- [1] D. P.-L. Philipp Rohlfshagen, Jialin Liu and S. M. Lucas, "Pac-Man Conquers Academia: Two Decades of Research Using a Classic Arcade Game," *IEEE Transactions on Computational Intelligence and AI in Games*, 2017.
- [2] S. Karakovskiy and J. Togelius, "The mario ai benchmark and competitions," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 1, pp. 55–67, 2012.
- [3] D. Loiacono, P. L. Lanzi, J. Togelius, E. Onieva, D. A. Pelta, M. V. Butz, T. D. Lonneker, L. Cardamone, D. Perez, Y. Sáez *et al.*, "The 2009 simulated car racing championship," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 2, no. 2, pp. 131–147, 2010.
- [4] M. G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling, "The arcade learning environment: An evaluation platform for general agents." *J. Artif. Intell. Res.(JAIR)*, vol. 47, pp. 253–279, 2013.
- [5] S. Ontanón, G. Synnaeve, A. Uriarte, F. Richoux, D. Churchill, and M. Preuss, "A survey of real-time strategy game ai research and competition in starcraft," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 5, no. 4, pp. 293–311, 2013.
- [6] D. Perez-Liebana, S. Samothrakakis, J. Togelius, S. M. Lucas, and T. Schaul, "General Video Game AI: Competition, Challenges and Opportunities," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 4335–4337.
- [7] M. Ebner, J. Levine, S. M. Lucas, T. Schaul, T. Thompson, and J. Togelius, "Towards a video game description language," 2013.
- [8] T. Schaul, "A video game description language for model-based or interactive learning," in *Computational Intelligence in Games (CIG), 2013 IEEE Conference on*. IEEE, 2013, pp. 1–8.
- [9] —, "An extensible description language for video games," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 6, no. 4, pp. 325–331, 2014.
- [10] J. Levine, C. B. Congdon, M. Ebner, G. Kendall, S. M. Lucas, R. Mikkilainen, T. Schaul, and T. Thompson, "General video game playing," *Dagstuhl Follow-Ups*, vol. 6, 2013.
- [11] D. Perez, S. Samothrakakis, J. Togelius, T. Schaul, S. Lucas, A. Couëtoux, J. Lee, C.-U. Lim, and T. Thompson, "The 2014 general video game playing competition," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 8, pp. 229–243, 2015.
- [12] R. D. Gaina, D. Perez-Liebana, and S. M. Lucas, "General Video Game for 2 Players: Framework and Competition," in *Proceedings of the IEEE Computer Science and Electronic Engineering Conference (CEECE)*, 2016.
- [13] R. D. Gaina, A. Couëtoux, D. J. Soemers, M. H. Winands, T. Vodopivec, F. Kirchgessner, J. Liu, S. M. Lucas, and D. Perez-Liebana, "The 2014 general video game playing competition," *IEEE Transactions on Computational Intelligence and AI in Games*, 2017.
- [14] A. Khalifa, D. Perez-Liebana, S. M. Lucas, and J. Togelius, "General video game level generation," in *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*. ACM, 2016.
- [15] A. Khalifa, M. C. Green, D. Pérez-Liébana, and J. Togelius, "General Video Game Rule Generation," in *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2017.
- [16] D. Pérez-Liébana, M. Stephenson, R. D. Gaina, J. Renz, and S. M. Lucas, "Introducing Real World Physics and Macro-Actions to General Video Game AI," in *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2017.
- [17] J. Liu, D. Perez-Liebana, and S. M. Lucas, "The Single-Player GVGAI Learning Framework - Technical Manual," 2017. [Online]. Available: http://www.liujialin.tech/publications/GVGAI_SingleLearning_manual.pdf
- [18] T. Joppen, M. Moneke, N. Schroder, C. Wirth, and J. Furnkranz, "Informed Hybrid Game Tree Search for General Video Game Playing," *IEEE Transactions on Computational Intelligence and AI in Games*, 2017.
- [19] D. Ashlock, D. Pérez-Liébana, and A. Saunders, "General Video Game Playing Escapes the No Free Lunch Theorem," in *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2017.
- [20] T. Geffner and H. Geffner, "Width-based planning for general video-game playing," in *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*, 2015.
- [21] D. Soemers, C. Sironi, T. Schuster, and M. Winands, "Enhancements for Real-Time Monte-Carlo Tree Search in General Video Game Playing," in *IEEE Conference on Computational Intelligence and Games (CIG)*, 2016.
- [22] G. A. Rummery and M. Niranjan, *On-line Q-learning using connectionist systems*. University of Cambridge, Department of Engineering, 1994, vol. 37.
- [23] J. Liu, "Gvgai single-player learning competition at ieeecig17," 2017. [Online]. Available: <https://www.slideshare.net/liujialin126/gvgai-singleplayer-learning-competition-at-ieee-cig17>
- [24] C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakakis, and S. Colton, "A survey of monte carlo tree search methods," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 1, pp. 1–43, 2012.
- [25] M. B. A. for General Video Games, "Torsten schuster," Master's thesis, Maastricht University, 2015.
- [26] D. J. Soemers, "Enhancements for real-time monte-carlo tree search in general video game playing," Master's thesis, Maastricht University, 2016.
- [27] D. Perez, S. Samothrakakis, and S. Lucas, "Knowledge-based fast evolutionary mcts for general video game playing," in *2014 IEEE Conference on Computational Intelligence and Games*. IEEE, 2014, pp. 1–8.
- [28] C. Y. Chu, T. Harada, and R. Thawonmas, "Biasing monte-carlo rollouts with potential field in general video game playing," p. 6p, 2015.
- [29] C. Y. Chu, H. Hashizume, Z. Guo, T. Harada, and R. Thawonmas, "Combining pathfinding algorithm with knowledge-based monte-carlo tree search in general video game playing," in *2015 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2015, pp. 523–529.
- [30] H. Park and K.-J. Kim, "Mcts with influence map for general video game playing," in *2015 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2015, pp. 534–535.
- [31] P. Auer, "Using confidence bounds for exploitation-exploration trade-offs," *Journal of Machine Learning Research*, vol. 3, no. Nov, pp. 397–422, 2002.
- [32] E. H. dos Santos and H. S. Bernardino, "Redundant action avoidance and non-defeat policy in the monte carlo tree search algorithm for general

- video game playing,” in *Proceedings do XVI Simpósio Brasileiro de Jogos e Entretenimento Digital*, 2017.
- [33] M. d. Waard, D. M. Roijers, and S. C. Bakkes, “Monte carlo tree search with options for general video game playing,” in *2016 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2016, pp. 47–54.
- [34] D. Perez-Liebana, S. Mostaghim, and S. M. Lucas, “Multi-objective tree search approaches for general video game playing,” in *Evolutionary Computation (CEC), 2016 IEEE Congress on*. IEEE, 2016, pp. 624–631.
- [35] A. Khalifa, M. Preuss, and J. Togelius, “Multi-objective adaptation of a parameterized gvgai agent towards several games,” in *International Conference on Evolutionary Multi-Criterion Optimization*. Springer, 2017, pp. 359–374.
- [36] I. Bravi, A. Khalifa, C. Holmgård, and J. Togelius, “Evolving uct alternatives for general video game playing,” in *The IJCAI-16 Workshop on General Game Playing*, 2016, p. 63.
- [37] A. Babadi, B. Omoomi, and G. Kendall, “Enhic: An enforced hill climbing based system for general game playing,” in *2015 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2015, pp. 193–199.
- [38] M. J. Nelson, “Investigating vanilla mcts scaling on the gvg-ai game corpus,” in *2016 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2016, pp. 402–408.
- [39] R. D. Gaina, J. Liu, S. M. Lucas, and D. Pérez-Liebana, “Analysis of vanilla rolling horizon evolution parameters in general video game playing,” in *European Conference on the Applications of Evolutionary Computation*. Springer, 2017, pp. 418–434.
- [40] B. Jia, M. Ebner, and C. Schack, “A gp-based video game player,” in *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*. ACM, 2015, pp. 1047–1053.
- [41] B. Jia and M. Ebner, “A strongly typed gp-based video game player,” in *2015 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2015, pp. 299–305.
- [42] R. D. Gaina, S. M. Lucas, and D. Pérez-Liebana, “Population Seeding Techniques for Rolling Horizon Evolution in General Video Game Playing,” in *2017 IEEE Conference on Evolutionary Computation (CEC)*. IEEE, 2017.
- [43] —, “Rolling Horizon Evolution Enhancements in General Video Game Playing,” in *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2017.
- [44] D. Perez Liebana, J. Dieskau, M. Hunermund, S. Mostaghim, and S. Lucas, “Open loop search for general video game playing,” in *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*. ACM, 2015, pp. 337–344.
- [45] H. Horn, V. Volz, D. Pérez-Liebana, and M. Preuss, “Mcts/ea hybrid gvgai players and game difficulty estimation,” in *Computational Intelligence in Games (CIG), 2013 IEEE Conference on*. IEEE, 2013, pp. 1–8.
- [46] J. van Eeden, “Analysing and improving the knowledge-based fast evolutionary mcts algorithm,” 2015.
- [47] C.-Y. Chu, S. Ito, T. Harada, and R. Thawonmas, “Position-based reinforcement learning biased mcts for general video game playing,” in *2016 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2016, pp. 444–451.
- [48] E. Ilhan, Erdem, and Ş. Uyar, “Monte Carlo Tree Search with Temporal-Difference Learning for General Video Game Playing,” in *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2017.
- [49] H. Van Seijen, A. R. Mahmood, P. M. Pilarski, M. C. Machado, and R. S. Sutton, “True online temporal-difference learning,” *Journal of Machine Learning Research*, vol. 17, no. 145, pp. 1–40, 2016.
- [50] I. Bravi, A. Khalifa, C. Holmgård, and J. Togelius, “Evolving game-specific ucb alternatives for general video game playing,” in *European Conference on the Applications of Evolutionary Computation*. Springer, 2017, pp. 393–406.
- [51] C. F. Sironi, J. Liu, D. Perez-Liebana, R. D. Gaina, I. Bravi, S. M. Lucas, and W. M. H. M. “Self-Adaptive MCTS for General Video Game Playing,” in *European Conference on the Applications of Evolutionary Computation*. Springer, 2018, p. to appear.
- [52] K. Kananusont, R. D. Gaina, J. Liu, D. Perez-Liebana, and S. M. Lucas, “The n-tuple bandit evolutionary algorithm for automatic game improvement,” in *Evolutionary Computation (CEC), 2017 IEEE Congress on*. IEEE, 2017, pp. 2201–2208.
- [53] B. Ross, “General video game playing with goal orientation,” Master’s thesis, University of Strathclyde, 2014.
- [54] P. Bontrager, A. Khalifa, A. Mendes, and J. Togelius, “Matching games and algorithms for general video game playing,” in *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2016.
- [55] A. Mendes, A. Nealen, and J. Togelius, “Hyperheuristic general video game playing,” *Proceedings of Computational Intelligence and Games (CIG)*. IEEE, 2016.
- [56] J. M. Gonzalez-Castro and D. Perez-Liebana, “Opponent Models Comparison for 2 Players in GVGAI Competitions,” in *Computer Science and Electronic Engineering Conference (CEEC), 2017 9th*. IEEE, 2017.
- [57] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998, vol. 1, no. 1.
- [58] S. Samothrakakis, D. Perez-Liebana, S. M. Lucas, and M. Fasli, “Neuroevolution for general video game playing,” in *2015 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2015, pp. 200–207.
- [59] A. Braylan and R. Miikkulainen, “Object-model transfer in the general video game domain,” in *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2016.
- [60] K. Kananusont, “General video game artificial intelligence: Learning from screen capture,” Master’s thesis, University of Essex, 2016.
- [61] K. Kananusont, S. M. Lucas, and D. Pérez-Liebana, “General Video Game AI: Learning from Screen Capture,” in *2017 IEEE Conference on Evolutionary Computation (CEC)*. IEEE, 2017.
- [62] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al., “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, p. 529, 2015.
- [63] S. O. Kimbrough, G. J. Koehler, M. Lu, and D. H. Wood, “On a feasible-infeasible two-population (fi-2pop) genetic algorithm for constrained optimization: Distance tracing and no free lunch,” *European Journal of Operational Research*, vol. 190, no. 2, pp. 310–327, 2008.
- [64] J. Nichols, “The use of genetic algorithms in automatic level generation,” Master’s thesis, University of Essex, 2016.
- [65] X. Neufeld, S. Mostaghim, and D. Perez-Liebana, “Procedural level generation with answer set programming for general video game playing,” in *Computer Science and Electronic Engineering Conference (CEEC), 2015 7th*. IEEE, 2015, pp. 207–212.
- [66] T. S. Nielsen, G. A. Barros, J. Togelius, and M. J. Nelson, “Towards generating arcade game rules with vgdI,” in *Computational Intelligence and Games (CIG), 2015 IEEE Conference on*. IEEE, 2015, pp. 185–192.
- [67] T. Machado, I. Bravi, Z. Wang, A. Nealen, and J. Togelius, “Shopping for game mechanics,” 2016.
- [68] T. Machado, A. Nealen, and J. Togelius, “Cicero: A mixed-initiative AI-assisted game design tool,” in *Proceedings of the 12th International Conference on the Foundations of Digital Games*, ser. FDG’17. New York, NY, USA: ACM, 2017.
- [69] —, “SeekWhence a Retrospective Analysis Tool for General Game Design,” in *Proceedings of the 12th International Conference on the Foundations of Digital Games*, ser. FDG ’17. New York, NY, USA: ACM, 2017, pp. 4:1–4:6. [Online]. Available: <http://doi.acm.org/10.1145/3102071.3102090>
- [70] J. Liu, J. Togelius, D. Pérez-Liebana, and S. M. Lucas, “Evolving game skill-depth using general video game ai agents,” in *2017 IEEE Conference on Evolutionary Computation (CEC)*. IEEE, 2017.
- [71] C. Guerrero-Romero, A. Louis, and D. Pérez-Liebana, “Beyond Playing to Win: Diversifying Heuristics for GVGAI,” in *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2017.
- [72] A. Khalifa, A. Isaksen, J. Togelius, and A. Nealen, “Modifying mcts for human-like general video game playing,” in *IJCAI*, 2016, pp. 2514–2520.
- [73] M. C. Green, A. Khalifa, G. A. Barros, and J. Togelius, ““Press Space To Fire”: Automatic Video Game Tutorial Generation,” in *Fourth Experimental AI in Games Workshop*, 2017.
- [74] D. Anderson, M. Stephenson, J. Togelius, C. Salge, J. Levine, and J. Renz, “Deceptive games,” *arXiv preprint arXiv:1802.00048*, 2018.
- [75] T. S. Nielsen, G. A. Barros, J. Togelius, and M. J. Nelson, “General video game evaluation using relative algorithm performance profiles,” in *European Conference on the Applications of Evolutionary Computation*. Springer, 2015, pp. 369–380.
- [76] S. M. Lucas, J. Liu, and D. Perez-Liebana, “The n-tuple bandit evolutionary algorithm for game agent optimisation,” *arXiv preprint arXiv:1802.05991*, 2018.
- [77] D. Pérez-Liebana, S. Samothrakakis, J. Togelius, T. Schaul, and S. M. Lucas, “Analyzing the robustness of general video game playing agents,” in *Proceedings of the IEEE Conference on Computational Intelligence and Games*, 2016.

- [78] X. Neufeld, "Procedural level generation with answer set programming for general video game playing," Master's thesis, University of Magdeburg, 2016.
- [79] R. D. Gaina, "The 2 player general video game playing competition," Master's thesis, University of Essex, 2016.
- [80] V. Volz, D. Ashlock, S. Colton, S. Dahlskog, J. Liu, S. Lucas, D. Perez-Liebana, and T. Thompson, "Gameplay evaluation measures," *Dagstuhl Follow-Ups*, 2017.
- [81] B. Pell, "Metagame in symmetric chess-like games," 1992.
- [82] A. Isaksen, D. Gopstein, J. Togelius, and A. Nealen, "Discovering unique game variants," in *Computational Creativity and Games Workshop at the 2015 International Conference on Computational Creativity*, 2015.
- [83] P. R. Williams, D. Perez-Liebana, and S. M. Lucas, "Ms. Pac-Man Versus Ghost Team CIG 2016 Competition," in *Computational Intelligence and Games (CIG), 2016 IEEE Conference on*, 2016.
- [84] P. Hingston, "The 2k botprize," in *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*. IEEE, 2009, pp. 1–1.
- [85] N. Shaker, J. Togelius, G. N. Yannakakis, L. Poovanna, V. S. Ethiraj, S. J. Johansson, R. G. Reynolds, L. K. Heether, T. Schumann, and M. Gallagher, "The turing test track of the 2012 mario ai championship: entries and evaluation," in *Computational Intelligence in Games (CIG), 2013 IEEE Conference on*. IEEE, 2013, pp. 1–8.
- [86] J. Lehman and R. Miikkulainen, "General video game playing as a benchmark for human-competitive ai," in *AAAI-15 Workshop on Beyond the Turing Test*, 2015.
- [87] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," *arXiv preprint arXiv:1606.01540*, 2016.
- [88] M. Johnson, K. Hofmann, T. Hutton, and D. Bignell, "The malmo platform for artificial intelligence experimentation." in *IJCAI*, 2016, pp. 4246–4247.
- [89] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton *et al.*, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, p. 354, 2017.