# Help File for GBG GUI

CONTENTS

[Arena & Arena Train Elements 1](#_Toc502746621)

[Agent Type 1](#_Toc502746622)

[Play 1](#_Toc502746623)

[InspectV 1](#_Toc502746624)

[Params X, Params O 1](#_Toc502746625)

[Train X 1](#_Toc502746626)

[Train O 1](#_Toc502746627)

[MultiTrain 1](#_Toc502746628)

[Agent menu 1](#_Toc502746629)

[Load Agent 1](#_Toc502746630)

[Save Agent 1](#_Toc502746631)

[Quick Evaluation 1](#_Toc502746632)

[Competition menu 1](#_Toc502746633)

[Single Compete 1](#_Toc502746634)

[Swap Compete 1](#_Toc502746635)

[Multi-Competition 1](#_Toc502746636)

[Param Tabs 1](#_Toc502746637)

[TD Params 1](#_Toc502746638)

[TD Params: Feature Sets 1](#_Toc502746639)

[NT Params 1](#_Toc502746640)

[MC Params 1](#_Toc502746641)

[MCTS & MCTSE Params 1](#_Toc502746642)

[MaxN Params 1](#_Toc502746643)

[Other Params 1](#_Toc502746644)

[Logs 1](#_Toc502746645)

[Options Menu 1](#_Toc502746646)

[Load Menu 1](#_Toc502746647)

[Available evaluators 1](#_Toc502746648)

[TicTacToe 1](#_Toc502746649)

[Hex 1](#_Toc502746650)

[2048 1](#_Toc502746651)

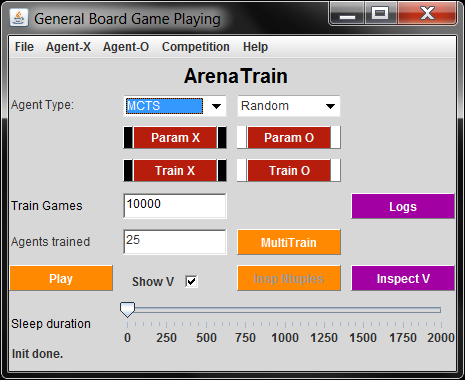
[Help 1](#_Toc502746652)

[Show Help File 1](#_Toc502746653)

[Help File in Browser 1](#_Toc502746654)

[Show TR-GBG.pdf 1](#_Toc502746655)

## Arena & Arena Train Elements



### Agent Type

The select boxes allow to select one of the following agent types for every player:

* **T****DS**: TD Learning agent according to [Sutton&Bonde 1993] with a linear net or backprop net (one hidden layer) as approximator for the value function, needs user-defined features (class TDAgent).
* **TD-Ntu****ple-2**: TD-Learning agent using N-tuples: a new version of TD-Ntuple, with afterstate and new lambda-mechanism [Jaskowski16] (class TDNTuple2Agt).
* **M****C**: Monte Carlo
* **MC****TS**: Monte Carlo Tree Search
* **MCTS Expectimax**: Monte Carlo Tree Search for nondeterministic games
* **Mi****nimax**: Tree search agent using Minimax strategy (w/o Alpha-Beta-search), only for 2-player games. Realizes perfect play for TicTacToe, but may ‘explode’ for more complex games.
* **M****ax-N**: the generalization of [Minimax](#Minimax) to N players for deterministic games.
* **Exp****ectimax-N**: the generalization of [Minimax](#Minimax) to N players for nondeterministic games.
* **Ra****ndom**: random playing agent
* **Hu****man**: human player

### Play

Play a game with the agents currently selected as “X” or “O” in the [Agent Type](#AgentType) combo boxes. If the currently selected agent (“X” or “O”) is not yet trained or the trained agent differs from the one in the combo boxes, an error message is displayed. If one of the selected agents is [Human](#Human), the user has to make the appropriate moves on the game board.

The starting player is always “X” (2-player games). The game starts from the initial start board (in many games the empty board). To start from other boards see [InspectV](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\InspectV).

If checkbox **Show V** is selected, the game board will display the game values for the agent who just made the **last** move. (Note the difference to [InspectV](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\InspectV), where the game values shown are those of the available moves for the **current** state.)

### InspectV

Inspect the value function V(**s’**) of the X-Player. The X-Player is

(a) before using the [Train X button](#Train_X): the agent in the X-agent select box (which is the first element of Types::GUI\_AGENT\_INITIAL, currently MCTS),   
(b) after using the [Train X button](#Train_X): the last trained X-Player.

Different states **s** can be set by the user in the action buttons in GameBoard:

* for **TicTacToe**: clicking a tile sets a black (X) or white (O) piece there,
* for **Hex**: clicking a tile sets a black (X) or white (O) piece there,
* for **2048**: clicking one of the four action buttons (up, right, …) performs this action, adds a new random tile and shows the values V(a’) of the then possible actions a’.

The value of V(**s’**) for each allowed successor **s’** of **s** is displayed on the GameBoard in a game-specific way.

When in INSPECTV mode, all other buttons (except [InspectV](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\InspectV) and [Play](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\Play)) are disabled.

* Another click on [InspectV](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\InspectV) ends the INSPECTV mode and enables again all other buttons.
* A click on [Play](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\Play) starts playing a game from the actual board, as configured by [InspectV](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\InspectV). This allows to test how an agent performs when it starts from a position different from the default initial position (e. g. empty board).

### Params X, Params O

Display the multi-tabbed Params window and set parameter in any of the tabs ([TD params](#TD_params) and so on, see below). The parameters are fetched from this multi-tabbed window when one of the train buttons, Train X, Train O, or MultiTrain is pressed. This is for the trainable agents. For the non-trainable agents, the parameters are fetched from this multi-tabbed window when Quick Eval, Compete, Multi Compete, or Save Agent (Arena menu) are issued or when buttons [Play](#Play) or [InspectV](#InspectV) are pressed.

### Train X

Fetch the parameter settings from the multi-tabbed Params window. Construct the X agent according to the X combo box and train it for “Train games” games.

During training: an Evaluator with mode [Quick Eval Mode](#QuickEvalMode) is called every [NumEval](#NumEval) training games and its result is shown in a JFreeChart XY-plot.

During training: If [StopTest](#StopTest)>0 and [StopEval](#StopEval)>0, the same Evaluator is called every StopTest training games. If the Evaluator signals “Training goal reached” (i.e. sufficient good play for a sufficient long period, see [Other params](#Other_params) for more details), the training is stopped prematurely.

After training, the trained player is evaluated (Evaluator.eval()). This is for example in the case TicTacToe:

“Success against random” = average success rate when playing 100 games against RandomPlayer, both as X and O (optimum: 1.0),  
“Success agains minimax” = success rate when playing an X- and an O-game against MinimaxPlayer (optimum: 0.0, i.e. always tie).

Note that the success rate in TicTacToe becomes negative, when the other player predominantly wins.

### Train O

Same as [Train X](#Train_X), but for the O-player.

### MultiTrain

Same as [Train X](#Train_X), but perform “Agents trained” training runs w/o plotting to the line chart (this is useful for running a training completely in background and to average performance over random fluctuations). Store all information needed to construct such a line chart later in agents[/<gameDir>](#gameDir)/csv/multiTrain.csv. (There are several R-script templates in directory resources/R\_plotTools for visualizing the results from one or several files multiTrain.csv.) Report the average success in console. Return the last trained agent as the agent for X.

## Agent menu

For each agent there is an Agent-\* menu with the following entries:

### Load Agent

### Save Agent

Agents are stored in and loaded from  
 agents/[<gameDir>](#gameDir)/<agentName>.agt.zip  
<gameDir> is either the name of the game or it is the name of the game plus a subdirectory characterizing the game subtype, e.g. Hex/04 in the case of a 4x4 Hex game.   
<agentName> is an arbitrary name characterizing the agent.

An agent is saved and loaded with **all** parameters that characterize it. On load, these parameters are automatically restored in the [Param Tabs](#ParamTabs).

### Quick Evaluation

Run an evaluation with the evaluator in mode [QuickEvalMode](#QuickEvalMode)

## Competition menu

This menu is not relevant for 1-player games.

### Single Compete

Make a competition “X vs O” consisting of “Games/Comp” games and report results.

### Swap Compete

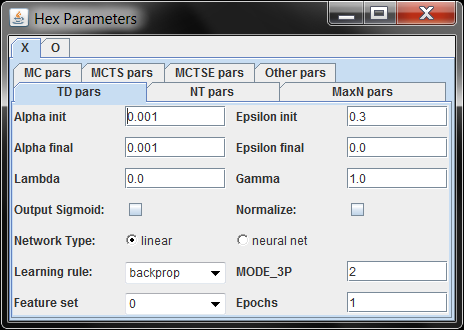
Swap the roles of X and O, i. e. make a competition “O vs X” consisting of “Games/Comp” games and report results.

### Multi-Competition

Perform multiple times competitions “X vs O” and report results. The agents (if trainable) are trained anew before each competition. The number of competitions is the value of “# competitions” in the sub-menu “Competition options”.

## Param Tabs

The parameters for all agents are set in the hierarchical multi-tabbed Params window:



Each player (here: “X” and “O”) has a multi-tabbed subwindow for setting the parameters for specific agents

### TD Params

Parameter for Temporal Difference Learning (for [TDS](#TDS) or [TD-Ntuple-2](#TD_Ntuple_2) agent):

* **Alpha init**: initial learn step size in first training episode.
* **Alpha final**: final learn step size in last training episode. The change from Alpha init to Alpha final is a geometric slope.
* **Epsilon init**: initial random move rate in first training episode.
* **Epsilon final**: final random move rate in last training episode. The change from Epsilon init to Epsilon final is a linear slope. A random move is done, if a random number is smaller than Epsilon. Epsilon may have values outside .
* **Lambda**: eligibility trace parameter . Usually, should hold.
* **Gam****ma**: discount factor in range [0, 1]
* **Network type** (for[TDS](#TDS) only):[linear] the output activation is either a linear function of the (generalized) input features or a backpropagation network with one hidden layer of size 15. For [TD-Ntuple-2](#TD_Ntuple_2), the network type is always linear.
* **Output sigmoid**:[without]should the output unit be with a sigmoid? If“with”, then the Fermi function  
    
  is used as sigmoid in case of agent type [TDS](#TDS).  
  If the agent is of type [TD-Ntuple-2](#TD_Ntuple_2), then there is either no output sigmoid or a Tangens hyperbolicus sigmoid:
* **Normalize**: Ifchecked, then each game score returned from the state observer is normalized to a range appropriate for the output sigmoid, that is range [0,1] in the case of Fermi function (TDS) and [-1,1] in the case of Tangens hyperbolicus ([TD-Ntuple-2](#TD_Ntuple_2)).
* **Learning rule** (for[TDS](#TDS) only): backprop or RPROP
* **MODE\_3P** (for [TD-Ntuple-2](#TD_Ntuple_2) only): Different update schemes to extend the normal TD update process to arbitrary N-player games
  + **0**: Update N independent value function V(s­t|p(i)) for all i=0,…,N-1 in each time step t. Works for arbitrary N, but weaker results for 2-player games than MODE\_3P=1 or =2 (unwanted crosstalk).
  + **1**: Update only for V(st|pt) and make an **N-ply** evaluation.
  + **2** [default]: Mixed version to take the best out of MODE\_3P=0 and =1:
    - **1-player games**: take MODE\_3P=0
    - **2-player games**: take MODE\_3P=1 with **1-ply** evaluation and use the symmetry V(st+1|pt) = – V(st+1|pt+1)
    - **N-player games with N>2**: take mode=0 (and live with the cross-talk)

If the value function is approximated by a neural network, the effective learning rate for the input-to-hidden weights is Alpha divided by the input-fan-in (size of input layer) and the learning rate for the hidden-to-output weights is Alpha divided by the hidden-fan-in (size of the hidden layer).

### TD Params: Feature Sets

Feature sets for [TDS](#TDS) player in the case of game **TicTacToe**:

* **0**: singlets/doublets/triplets for “self” and “opponent”
* **1**: singlets/doublets/triplets for “X” and “O”
* **2**: singlets/doublets + diversity + crosspoints for “X” and “O”
* **3**: same as **2** + the 9 “raw” board positions
* **4**: same as **2** + occupation midpoint, occupation corner
* **5**: same as **2** + …
* **9**: the 9 “raw” board positions

### NT Params

Parameter for n-Tuple TD-Learning (for [TD-Ntuple-2](#TD_Ntuple_2) agent):

* **TC**: Temporal Coherence on/off
* **INIT**: (only if TC) TC train counters are initialized with random numbers from [0,INIT]
* **TC factor**: (only if TC) [Immediate] update of TC factors or [Accumulting] update after N episodes
* **Episodes**: (only if TC factor==Accumulating) the number N of episodes
* **nTuple randomness**: =true: use fixed n-tuples, =false: construct random n-tuples
* **nTuple generation**: RandomWalk (connected n-tuples), RandomPoints (arbitrary points on the board)
* **# of nTuples**
* **nTuple size**
* **USESYMMETRY**:=true: during training and retrieval each board state activates also all equivalent board states (those connected by the symmetries of the game, e.g. four rotations\*two reflections in the case of 2048), =false: do not use the equivalent board states
* **AFTERSTATE** (only for nondeterministic games, e.g. 2048): Ifchecked, use the afterstate logic as described in [Jaskowski16]: The argument s’ in V(s’) is the state after an action, **prior** to adding the non-deterministic element. If not checked, the argument s for V(s) is the next state (including the non-deterministic element).

### MC Params

Parameter for [MC](#MC) agent (class MCAgent):

* **Iterations**: [1000] how many rollouts are performed
* **Rollout Depth**: [20] the maximum rollout depth (how many plys)
* **Number agents**: [1] use an ensemble of this number of MC agents and base the action decision on majority vote

### MCTS & MCTSE Params

Parameter for [MCTS](#MCTS) agent (class MCTSAgentT):

* **Iterations**: [1000] how many rollouts are performed
* **K[UCT]**: [sqrt(2)] balances exploitation and exploration in the UCT formula
* **Tree Depth**: [10] the maximum MCTS tree depth
* **Rollout Depth**: [10] the maximum rollout depth (how many plys)

MCTSE has the same parameters plus **Max Nodes**, the maximum number of allowed nodes (Expectimax nodes) in the tree

### MaxN Params

Parameter for [Max-N](#MaxN), [Expectimax-N](#ExpectimaxN) or [Minimax](#Minimax) agents:

* **Tree Depth**: [10] the maximum tree depth (recursion depth) of the agent.
* **Minimax Hash**: [true] use hash map to store already visited states (only Minimax). MaxN and ExpectimaxN have no hash map

### Other Params

During or after training an agent, this agent can be evaluated by an [evaluator](#evaluators) (see below). If such an evaluator signals “success” (for a long enough training period of [StopEval](#StopEval) training games), then training might be stopped prematurely.

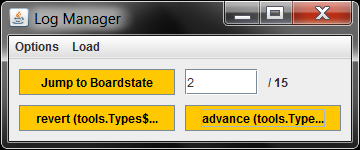
Settings:

* **Qui****ck Eval Mode**: An evaluator with this mode is used in [Quick Evaluation](#QuickEvaluation), in training and in multi-training. This is the evaluator with the [StopEval](#StopEval) test to end training prematurely. This is as well the evaluator whose performance during training is shown in the JFreeChart XY-plot every [NumEval](#NumEval) training games.
* **Train Eval Mode**: If different from [Quick Eval Mode](#QuickEvalMode), another evaluator is constructed in training and multi-training. It is evaluated in parallel to assess the strength of an agent after training from a different perspective. If Train Eval Mode is identical to [Quick Eval Mode](#QuickEvalMode), no second evaluator is constructed.
* **N****umEval**: [100] after every NumEval training games the performance of the trained agent is evaluated (success against [Minimax](#Minimax), success against [Random](#Random)) and the success against [Minimax](#Minimax) is plotted in a JFreeChart window. Choose higher values for NumEval to speed up training.
* **Sto****pTest**: [0] after every StopTest training games, the [Quick Eval Mode](#QuickEvalMode) evaluator is called to see if we can stop training prematurely. If 0, this Evaluator is never called and so training is never stopped prematurely.
* **Sto****pEval**: [0] number of *consecutive* games that the [Quick Eval Mode](#QuickEvalMode) evaluator has to signal “success” before training is stopped prematurely. If 0, training is never stopped prematurely.
* **Choose Start 01**: If not checked, each training episode is started from the default (empty) game board. If checked, it is in 50% the empty game board and the other 50% are distributed equally on one of the possible 1-ply successors of the empty board. Increases the exploration.
* **Learn from RM**: If not checked, do not learn during training when a random move occurs. If checked, learn from each move (random action or not).
* **Reward = Score**: The boolean rewardIsGameScore is passed to StateObservation’s getReward(rewardIsGameScore) and. Depending on the boolean value, getReward returns either the usual game score or another (game-specific coded) reward.
* **Wrapper nPly**: during [Play](#Play), [Quick Evaluation](#QuickEvaluation), [Competition](#_Competition_menu) or [InspectV](#_InspectV): use either [Max-N](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\MaxN) or [Expectimax-N](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\ExpectimaxN) with this nPly as a wrapper around the actual agent. Which of both wrappers is used depends on the game: [Max-N](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\MaxN) for deterministic, [Expectimax-N](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\ExpectimaxN) for nondeterministic games. The wrapper constructs a tree of depth nPly and once the tree leaves are reached, it asks the wrapped agent for an game value estimate. This increases the performance in most cases, but be cautious with nPly>4: Depending on the branching factor of the game, this may lead to exploding computation times.

## Logs

During game play, played episodes can be logged. Logged episodes can be restored and re-played. Episodes are stored in   
 logs/[<gameDir>](#gameDir)/<game>\_<date>\_<time>.gamelog  
See here for explanation of [<gameDir>](#gameDir).

Button **Logs** brings you to the Log Manager Window:



### Options Menu

* **Logging enabled**: [true] if checked, log every episode started via [Play](#Play)
* **Advanced logging**: [true] if checked, enable advanced logging, i.e. the possibility to restore from a temporary gamelog when the main program has crashed before the logging session was closed normally, see [LogManager-QQ2.pdf](file:///C:\Users\wolfgang\Documents\GitHub\GBG\resources\LogManager-QQ2.pdf) (in German) for details.
* **Verbose**: [true]
* Compile temporary gamelog
* Delete temporary gamelogs

### Load Menu

**Load Gamelog** opens a file chooser for logs/[<gameDir>](#gameDir). After selecting a .gamelog file, the episode stored in that file is loaded and can be replayed via buttons **advance** and **revert**, or you enter a board state number and select it with **Jump to Boardstate**. In either case, the gameboard window will display the selected state of that episode.

## Available evaluators

Each game has an Evaluator class which can construct evaluators in different modes.

### TicTacToe

* mode 0: competition (100 games) against RandomAgent
* mode 1: competition (1 game) against Minimax
* mode 2: competition (10 games) against Minimax from 10 different start states
* mode 9: measure rate of correct decisions and game function delta on a set of 24 states

An evaluation of EvaluatorTTT is termed a “success”, if its return value is above the threshold m\_thresh (currently m\_thresh = {0.8,-0.15,-0.15, 0.85} for the different modes in source code).

### Hex

* mode -1: disable evaluation.
* mode 0: competition (3 episodes, HexConfig.EVAL\_NUMEPISODES) against MCTS as 2nd player, always starting from empty board. MCTS uses 10N-1 iterations where N= min(HexConfig.BOARD\_SIZE,5) is the board size, treeDepth=10, and rolloutDepth=200. Strong evaluator for board sizes up to and including 5x5. No guarantees for 6x6 or higher. Tends to require a lot of memory for 7x7 and up.
* mode 1: competition (100 episodes) against Random as 2nd player. Very weak but fast evaluator.
* mode 2: competition (100 episodes) against Minimax as 2nd player. Strong evaluator if applicable. But **only applicable for up to 4x4** board sizes (otherwise memory & time requirements too large), and even for 4x4 boards it takes quite a while on first pass.
* mode 10: perform one competition consisting of 3 (HexConfig.EVAL\_NUMEPISODES) episodes, for each different start state coded in HexConfig.EVAL\_START\_ACTION[N][]. All start states are winning boards. The agent to be evaluated makes the next move against **MCTS as opponent**. MCTS uses 10K-1 iterations where K = min(HexConfig.BOARD\_SIZE,5) is the board size, treeDepth=10, and rolloutDepth=200. More statistically reliable than the mode-0 evaluator, but the remarks there about MCTS apply as well. The mode-10 computation time is K times longer than mode-0 evaluator (K = length(HexConfig.EVAL\_START\_ACTION[N][])).
* mode 11: perform one competition consisting of 3 (HexConfig.EVAL\_NUMEPISODES) episodes, for each different start state coded in HexConfig.EVAL\_START\_ACTION[N][]. All start states are winning boards. The agent to be evaluated makes the next move against the agent loaded from **TDReferee.agt.zip** **as opponent** (if this file is available in agents/[<gameDir>](#gameDir)).

### 

* mode 0: use Evaluator2048: average score from 50 (ConfigEvaluator.NUMBEREVALUATIONS) evaluation episodes
* mode 1: use Evaluator2048\_BoardPositions: load a large set of board positions from file gameStates.ser, analyze them when applying MC and MCTSE agents to them. Is the agent to be evaluated used at all??
* mode 2: use Evaluator2048\_EA: perform a CMA-evaluation of the heuristics (see [Kutsch2017])

## Help

### Show Help File

Toggle the display of this help text in a HTML window.

### Help File in Browser

The same in a browser window.

### Show TR-GBG.pdf

Show the document “**The GBG Class Interface Tutorial: General Board Game Playing and Learning**” by Wolfgang Konen.