

1 We thank all the reviewers for the constructive comments. We address the major concerns below.

2 **Q1. Reproducibility:** 1) learning to draft details; 2) feature details; 3) discussions on the computing resources used.

3 **1)** When drafting, a search tree is built iteratively, with each node representing a state and each edge representing the  
4 action (pick a hero) resulting to a next state. The search tree is updated based on four steps of MCTS. The selection,  
5 expansion and backup steps are the same as the normal MCTS. For the prediction step, we use a value network to predict  
6 the value of the current state. After 1600 iterations, we pick heroes based on the child of the root node with the most  
7 visits. For the win-rate predictor and value network, the dimensions of 3 FC layers are [200,128],[128\*2,128],[128,1]  
8 and [200,128],[128\*2+1,128],[128,1], respectively. The learning rate is set to 0.001 with Adam. The MCTS drafting  
9 match dataset contains about 30 million matchups, covering all the heroes in our hero pool uniformly. The source code  
10 for drafting will be released. **2)** We have added more details on the feature design, including the range of features,  
11 feature dimension, the way of normalization, and how these features were developed. **3)** We believe that using a large  
12 amount of training resource is not a weakness of the proposed method, considering the nature of the problem we solve.  
13 First, the computing resources required for complex game-playing AI programs are often non-trivial, e.g., AlphaGo  
14 (280 GPUs), OpenAI Five (1920 GPUs), AlphaStar (3072 TPUv3 cores), and this work (320 GPUs). Second, we will  
15 publish our infrastructure (in progress). Third, we will develop subtasks of MOBA-game-playing for the AI community.

16 **Q2. Novelty:** the building blocks have been reported in the literature, e.g., curriculum learning, MCTS, etc.

17 The way we use the building blocks and the problem we solve are both new, which has been clearly pointed out  
18 by R1, R3 and R4. This work’s focus is on an unsolved MOBA-game-playing problem, rather than the standalone  
19 building blocks themselves. Individually, each of the employed techniques has been innovatively adapted. For example,  
20 curriculum learning is just a conceptual framework; how to design a curriculum for MOBA AI remains unexplored.

21 **Q3. Fairness concerns:** 1) AI’s reaction compared to human player; 2) use of full information in the value network.

22 **1)** MOBA is different from general RTS. In RTS like StarCraft, a player controls many game units, and one can always  
23 choose to switch to other game units whose skills and attacks are usable. In this case, it is a must to restrict the EPMs.  
24 By comparison, in MOBA, one player controls one hero. All the hero skills have cooldown limit and the normal attack  
25 has attack-frequency-limit. In this case, reaction-time is generally used as the measurement for fairness in MOBA.  
26 OpenAI Five has a discussion on this (see Section 4.1 in their paper). “In Dota 2, the key measure of human dexterity is  
27 reaction time, contrast with RTS games like StarCraft...” **2)** Please be assured that this will not cause any unfairness.  
28 In our model, we ONLY use the full information in the value network. As we mentioned in Section 3.2–Value Updates:  
29 “Note that this is performed only during training, as we only use the policy network during evaluation.”

30 **Q4. Presentation & organization:** 1) the paper is organized in a way that reads like a list of design choices and features;  
31 2) more comparisons to a recent MOBA AI work by Ye et al. [39]; 3) removing the non-RTS part in related work.

32 **1)** As suggested by R4, we will improve the presentation in two ways: 1) further emphasizing the newly proposed parts,  
33 e.g., CSPL, MHV, and drafting; 2) re-organizing the sections in Method in an easier-to-read fashion. **2)** This work and  
34 [39] have completely different focuses. [39] studies MOBA 1v1 game. while this paper studies the scalability-related  
35 issues when expanding the hero pool in MOBA 5v5 game. The similarities between this work and [39] include: the  
36 modeling of action heads (the value heads are different); off-policy correction; the way of conducting human evaluation.  
37 We will make it clearer for a better read. **3)** We will remove the non-RTS part in related work.

38 **Q5. CSPL related:** 1) the rule of advancing to next curriculum; 2) the non-CSPL experiments haven’t been run long  
39 enough to make a clear call that they won’t achieve similar ELO values; 3) in Fig 3., Phase 1 and Phase 2 experiments  
40 need to be run for 192/480h to see what a difference CSPL actually makes.

41 **1)** The rule of advancing to the next curriculum in CSPL is based on the convergence of Elo scores. The convergence in  
42 student-driven policy distillation is not sensitive to the hero pool size. Both 20-hero and 40-hero cases can converge  
43 rapidly within two days. **2)** First, Fig. 3-d aims to show that CSPL can achieve faster learning than the non-CSPL  
44 counterpart (we can already observe CSPL speeds it up). Second, we kept running it for another 3 days; still, the  
45 non-CSPL did not achieve a similar Elo score. **3)** Phase 1 and Phase 2 are parts of the CSPL method itself.

46 **Q6. Various minor issues:** 1) Computation time in petaflop/s; 2) #heroes in OpenAI Five; 3) Minimax or AlphaBeta  
47 search, why computationally infeasible; 4) How many players in top 0.04%; 5) Why introduce multiple value heads; 6)  
48 Why the observation delays so long. 7) Elo of the never-acting bot; 8) #games played between the different picking  
49 strategies? 9) Unclear how LSTM can tackle imperfect information. 10) Grammar, phrasing and reference problems.

50 **1)** 1.523 PFlops/s-days. **2)** 17. **3)** Minimax is used in OpenAI Five; AlphaBeta pruning is not vanilla Minimax;  
51 because of the 213,610,453,056 combinations in a 40-hero case. **4)** 120,000. **5)** For better value estimation, as stated.  
52 **6)** To make the AI’s reaction speed comparable to human players. **7)** 0. **8)** 1000. **9)** we meant to say that it can help  
53 infer unobservable information. **10)** We have done a thorough proofreading.