Exploring the Hearthstone Deck Space

In:*Proceedings of the Foundations of Digital Games 2018*

Aditya Bhatt
Tandon School of Engineering
New York University
Brooklyn, New York, 11201
apb462@nyu.edu

Scott Lee
Independent
Hillsborough, New Jersey, 08844
randomperson2727@gmail.com

Fernando de Mesentier Silva
Tandon School of Engineering
New York University
Brooklyn, New York, 11201
fernandomsilva@nyu.edu

Connor W. Watson
Ying Wu College of Computing
New Jersey Institute of Technology
Newark, New Jersey, 07103
cww5@njit.edu

Julian Togelius
Tandon School of Engineering
New York University
Brooklyn, New York, 11201
julian@togelius.com

Amy K. Hoover
Ying Wu College of Computing
New Jersey Institute of Technology
Newark, New Jersey, 07103
ahoover@njit.edu

ABSTRACT

A significant issue in game balancing is understanding the game itself. For simple games end-to-end optimization approaches can help explore the game's design space, but for more complex games it is necessary to isolate and explore its parts. Hearthstone, Blizzard's popular two-player turn-taking adversarial card game, has two distinct game-playing challenges: choosing when and how to play cards, and selecting which cards a player can access during the game (deckbuilding). Focusing on deckbuilding, four experiments are conducted to computationally explore the design of Hearthstone. They address the difficulty of constructing good decks, the specificity and generality of decks, and the transitivity of decks. Results suggest it is possible to find decks with an Evolution Strategy (ES) that convincingly beat other decks available in the game, but that they also exhibit some generality (i.e. they perform well against unknown decks). Interestingly, a second ES experiment is performed where decks are evolved against opponents playing the originally evolved decks. Since the originally evolved decks beat the starter decks, and the twice evolved decks beat the originally evolved decks, some degree of transitivity of the deck space is shown. While only a preliminary study with restrictive conditions, this paper paves the way for future work computationally identifying properties of cards important for different gameplay strategies and helping players build decks to fit their personal playstyles without the need for in-depth domain knowledge.

CCS CONCEPTS

• **Applied computing → Computer games**; • **Mathematics of computing → Evolutionary algorithms**;

FDG18, August 7–10, 2018, Malmö, Sweden

ACM ISBN 978-1-4503-6571-0/18/08…$15.00
https://doi.org/10.1145/3235765.3235791

KEYWORDS

Hearthstone, Game Balancing, Deck Building, Evolutionary Computation, Evolution Strategies

ACM Reference Format:
Aditya Bhatt, Scott Lee, Fernando de Mesentier Silva, Connor W. Watson, Julian Togelius, and Amy K. Hoover. 2018. Exploring the Hearthstone Deck Space In:*Proceedings of the Foundations of Digital Games 2018.* In *Foundations of Digital Games 2018 (FDG18), August 7–10, 2018, Malmö, Sweden.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3235765.3235791

1 INTRODUCTION

Before fine-tuning a game, a designer must first understand how the different components in the game systems behave. However, because a game is typically a complex system of many interlocking parts, understanding its many intricacies and quirks is often even challenging for its creator. In part these complexities are why the game design and development process usually incorporates multiple rounds of playtesting. Typically imbalances are most obvious through these iterations with multiple playtesters.

However, one way to understand these imbalances is through artificial intelligence based testing the mechanics and dynamics. Typically these AI algorithms are designed to automatically play some part of the game, and then vary some aspect of the game and observe the resulting changes to output. Based on this variation in outcome an optimization algorithm can be used to find game variants that satisfy some criteria. At its simplest, one could isolate one or a small handful of continuous values, and optimize those values so that the game achieves parity in winrate between agents. This is often referred to as game balancing or game tuning [24]. But approaches have been pursued that identify a larger number of game dimensions, demonstrating that search in a multi-dimensional game space can be used to find games that are distinctly different from each other but still playable [12].

For more complex games, end-to-end optimization approaches are infeasible. Instead, specific questions about the game's mechanics or dynamics can be asked, and agent-based playthrough can be used to investigate particular aspects or subsystems of the game. For example, agents implementing different playstyles can help investigate how a game, level, or map affords, rewards, or punishes different styles and strategies [5, 11, 17].

Hearthstone, a popular two-player competitive digital collectible card game published by Blizzard [7], is an example of a game that is too complex for end-to-end optimization, but requires regular re-balancing to ensure quality. In Hearthstone players initially choose one of nine different *heroes*, which will determine the types of playable cards and special abilities, ultimately influencing play style. They then build decks, from which cards are drawn during play. The objective is to reduce the opponent hero's health to zero through minions and spells, which are invoked by playing cards. Broadly, playing Hearthstone breaks down into building a deck and a corresponding strategy in advance of a game, and executing upon said strategy during a game. While not unrelated, these skills can to some degree be seen as separate.

In this paper we conduct several experiments to computationally explore the design of Hearthstone, in particular the difficulty of constructing good decks, the specificity and generality of decks, and the transitivity of decks. We use a standard game-playing agent with fixed strategy parameters to focus the investigation on the decks. First, we investigate the difference in performance of the same agent over multiple random decks when playing against three standard *starter* decks, which are the Blizzard-made decks available to any player who has completed the tutorial. We then evolve decks to perform as well as possible against these fixed decks, and find that it is in all cases relatively straightforward to do so. We then investigate the robustness of the evolved decks when changing the strategy parameters of the game-playing agent, in particular changing from an *aggro* or aggressive to a *control* style of play. We also look at how the evolved decks perform against decks different from the one they were evolved against. Finally, we conduct another set of experiments where we evolve new decks specifically to beat the evolved decks, in order to investigate whether the strength of the deck is transitive, (i.e. if deck B wins over deck A, and deck C wins over deck B, will deck C also win over deck A?)

2 BACKGROUND

The idea of testing games automatically through algorithmic playthrough is an old one. In the search-based procedural content generation paradigm, simulation-based testing refers to testing some aspect of a game (such as a level) with game-playing agents [23]. Testing can indicate whether the level is playable to whether there is sufficient *skill depth* in the sense that agents of different performance achieve different results [19]. However, a more immediately practical case for automatic game testing is tuning games, i.e. automatically making small changes to the game to improve some measure such as balance or player progression.

A number of recent research papers address aspects of game tuning. One of the earliest studies by Hom and Marks [10] proposes designing balanced board game by altering the game rules with AI techniques. Jaffe et al. [13] introduces a method for measuring aspects of game balance [13], and shows how the techniques can facilitate quick progress in balancing an educational card game. Volz et al. [24] introduces another balancing method for decks of the Top Trumps game. Here the authors compare the processes of automatic and manual balancing and assess the quality of computer balanced decks in relation to existing ones. Mahlmann et al. [18] analyze three different agents who play and select cards, aiming to make a balanced deck in the game Dominion. The authors show that specific cards can lead the game to balance, independent of player skill level. The decks evolved in this work may point to interesting features related to the game's balance, such as quality of certain cards and characteristics of the card set pertaining to different character classes. Similarly, Krucher [16] introduces AI-agents that play and modify cards in a collectible card game to balance the pool of available cards.

Automated game balancing can also help balance modern board games. de Mesentier Silva et al. [5] presents methods to automate stages of the playtesting process with agents. By having multiple agents play the game Ticket to Ride, the authors can extract data that could help designers make informed decisions about game tuning. A rather different approach is that of Dormans [6] who presents a framework to balance games through diagramming the flow of resources in a game.

2.1 Playing Collectible Card Games

Because of the large amount of hidden information (the opponent's hand and deck), stochasticity in starting conditions (each player's deck is shuffled), and a medium-to-high branching factor (a varying number of cards can be played, in different order), it is difficult to develop good game-playing algorithms for Hearthstone and related collectible card games, such as *Magic: The Gathering*. Standard tree search methods struggle not only with the branching factor but also with searching into the opponent's turn, as in general very little is known about the opponent's hand or deck.

Some approaches focus on mitigating the large amount of hidden information by learning the value of game states. With neural networks, Jakubik [14] learn state values with sparse autoencoders. The Advances in Artificial Intelligence and Applications even offers a data mining challenge, where participants devise methods for predicting the value of game states [15].

Others focus on many iterations of automatic playthroughs rather than learning the value of game states, most though adapting Monte Carlo Tree Search (MCTS) to work in Hearthstone. Santos et al. [20] present special purpose heuristics for MCTS to achieve competent play while Zhang and Buro [25] decrease the branching factor by modifying MCTS to bucket similar nodes and training policy networks to guide the rollouts. Tree search-based approaches can also facilitate simulation-based testing; [26] test a Hearthstone-like game called Cardonomicon through simulations. While these approaches all search at the level of individual Hearthstone actions, other approaches aim to build a symbolic representation of the Hearthstone space and reason in that space [22].

Two studies address the role of decks in collectible card games. Bjørke and Fludal [4] evolved decks for Magic: The Gathering, a game with some strong similarities to Hearthstone but also some important differences. The previous work most closely related to the current paper is [9] where, decks are evolved and evaluated based on how much they win over a set of decks. In other words, decks are evolved for general playing strength. No subsequent analysis was done on the evolved decks.

3 HEARTHSTONE

Hearthstone is a turn-based, two-player, adversarial, collectible card game published by Blizzard with over 70 million players (as of March 2017)[7]. In this free-to-play, digital game, players collect cards to compose decks that are then played against others. Players acquire these cards most often through purchasing packs of five randomly selected cards. At the time of publication, Hearthstone has over 1400 unique cards of five rarities: Free, Common, Rare, Epic, and Legendary, where rarest cards are perceived as the most powerful and hardest to acquire. While at first glance Hearthstone gameplay may be perceived primarily as about which cards to play and when, as players advance, a large part of ensuring successful matches lies in deckbuilding (i.e. choosing which 30 cards to play out of the over 1400 possibilities). Gameplay then can be split into two main parts: building decks and playing matches, where players must first build a deck before battling an opponent.

To build a deck players must first choose one of nine possible heroes available in the game: Druid, Mage, Hunter, Paladin, Priest, Rogue, Shaman, Warlock, and Warrior. While each type of hero is represented by a unique avatar that the player assumes, it also provides access to a unique hero ability and set of exclusive, hero-specific cards that are a subset of the entire set of available cards. Players then choose exactly 30 cards to compose a deck from the hero-neutral and hero-specific card sets.

After creating a deck, players are then matched with others for one versus one play. Each hero starts with 30 health, and to win a player must reduce their opponent's health to zero. To reduce an opponent's health and protect themselves, players draw random cards from their deck to place in their hand, not visible by the opponent. Cards from a hand are then available to place on the common board with the goal of reducing the opponent's health.

Mana is the main resource of the game, and each card has a cost to place on the common board (i.e. mana cost). Players' collections of mana (i.e. mana pool) are privately held and reset at the beginning of each turn. Each turn allows access to an additional mana unit, where players start with one mana and end with up to ten. Cards costing more mana are often more powerful than those costing less. However, players can be defeated before being able to access their more powerful cards in later turns.

Types of cards can be subdivided into two functionalities, spells and minions. Spell cards are instantaneous once cast and are not inherently represented by units on the board, while minion cards are units that remain on the board once played. Spells are often triggered once and then disappear, while minions remain indefinitely in-play. Minions have two attributes (i.e. health and attack power) and disappear only when their health is depleted through attacks by the opponent's hero or other minions. When minions attack each other or heroes, they lose health points equal to the others attack power and are destroyed when their health is reduced to zero. When a minion's health is reduced to zero, it is destroyed and removed from the board. When minions attack the opponent's hero, the opponent loses health points equal to that Minions attack power. Most often minions cannot perform attacks on the same turn they were summoned, or placed on the common board. Minions often have additional attributes, such as *taunt* (i.e. the opponent must destroy this minion before attacking the hero), *battlecry* (i.e. an effect that occurs when the minion is first played), and *charge* (i.e. the minion can attack on the same turn it is played). Both spell and minion cards cost mana to play, although a few cost zero.

In Hearthstone players alternate turns, which can consist of playing one or more cards and attacking with one or more of their minions. At the start of each turn, players draw the top card of their deck, which has been shuffled and placed face down at the beginning of the game. The moment a player's health is reduced to zero, the match is over and the player with the most health wins.

4 APPROACH

To computationally explore the design of Hearthstone and the role of deck building, different combinations of cards are composed into decks by an evolution strategy (ES) algorithm, a process that optimizes the decks based on the principles of natural evolution. Simulated Hearthstone games are played in a Hearthstone simulator called SabberStone[8], which enables the evaluation of decks.

4.1 Building Competitive Decks with an Evolution Strategy

While there were approximately 382 cards when Hearthstone was originally released, Blizzard regularly adds new cards to increase replayability and balance the game resulting in over 1400 at present. In part because there are nine heroes each with a unique hero ability and hero-specific cards, balancing gameplay is nontrivial. While it would be valuable to explore balance and deck building based on all the cards in Hearthstone, experiments in this paper focus on those that beginning players can access, called the basic set [2]. There are 43 cards in this set that any hero can play, and each hero can play 10 cards unique to their class. These cards are unlikely to change with future expansions and are less complex than some cards developed later in the game.

Because it is unknown a priori which cards comprise a competitive deck and the inherent challenges of determining the value of particular cards, the search for decks in this paper is performed through a type of evolutionary algorithm called an evolution strategy [21]. The ES tries to optimize the decks by adding and removing cards and testing how well they do in games against an opponent. The idea is that cards important to a deck's success are preserved through evolution, while mutation helps discover new cards that can increase a deck's winrate. Because it is expected that Blizzard has fine-tuned starting decks to some basic level of proficiency, decks are evolved against games with an the opponent playing the *starter deck*, that is the deck composed of basic cards that are selected by Blizzard for new players during the tutorial.

4.2 Hearthstone Simulation and AI Agents

SabberStone is a Hearthstone simulator developed by the HearthSim community [8], a community of programmers who develop tools for Hearthstone players. SabberStone has a player-facing GUI, and natively supports AI-based agents who implement the five different meta-level strategies discovered by human players over time: aggro, control, fatigue, midrange, and ramp. Strategies differ primarily in whether they prioritize large health differences between heroes or board control, and whether they often result in early, mid, or late-game victories. Most strategies are based on some combination

```
procedure AggroHeuristic(state)
    if GameOver then
        f(state) = WIN ? MaxValue : MinValue
    else
        if Player has Minions and Opponent has None then
            f(state) += 1000
        end if
        if Opponent has Minions with Taunt then
            f(state) -= 1000×(Sum Health of Minions with Taunt)
        end if
        f(state) += (Sum of Atk of Player Minions)
        f(state) += 1000 * (Player Health - Opponent Health)
    end if
    return f(state)
end procedure
```

Figure 1: *Aggro Strategy.* Each agent in SabberStone selects actions based on one of five heuristics: aggro, control, midrange, fatigue, and ramp. Experiments in this paper implement the aggro strategy, which optimizes based on a large health difference between heroes and the number and type of minions on the board.

of a control heuristic (i.e. controlling the board until cards with high mana costs can be played) or aggro heuristic, which prioritizes maintaining minions on the board and aggressive play.

In the experiments in this paper, agents play Hearthstone with a greedy search algorithm, where each node is a game state. Search is performed with a fixed number of nodes per level (4 nodes) and maximum depth (13 nodes). Decisions are turn-local, meaning that they only calculate what the best actions are for a given turn rather than considering how those decisions may affect future turns. While future work may consider more sophisticated behavior, gameplay with this agent is still computationally expensive considering that a variable number of actions can be taken in a single Hearthstone turn and the order can affect the outcome.

While there are five tree-search strategies to choose from in SabberStone and many more potential strategies yet undiscovered, the aggro or aggressive strategy is one of the more basic strategies expected to perform well with the turn-local agent. Because it also appears similar to the logic humans follow when playing an aggro deck, agents for both the player and the opponent make gameplay decisions with SabberStone's implementation of an aggro strategy (described in figure 1). Compared to the other strategies, aggro primarily optimizes toward a large health difference between heroes, and punishes a board state when the opponent has minions with taunt in play. Minions with taunt implicitly defend a hero by forcing the opponent to attack it before the hero. Aggro also rewards states where one player has minions on the board and the other does not and when minions have more attack power and health, but prioritizes health differences between heroes. Future work will investigate how this strategy may affect deckbuilding by comparing it to human play.

5 EXPERIMENTS AND RESULTS OVERVIEW

To explore automated deck building and the properties of these evolved decks, four sets of experiments are proposed. First in section 6, decks are evolved for the hunter, paladin, and warlock heroes, whose unique hero abilities and class cards enable humans to play them with successful aggro strategies. Not only is the idea to show that successful decks can be evolved, but also explore commonalities, hinting toward a schema for successful decks of basic cards.

To explore the specificity of the decks, in section 7 each of the evolved decks from the previous experiment play $20,000$ games with each of the nine heroes, 10,000 against an opponent playing a starter deck, and 10,000 against an opponent playing the community-made deck of basic cards that are considered more powerful [1]. While the evolved decks are optimized to beat heroes of the same class, the experiments look at the winrates against all nine heroes. For further validation, and to observe general trends these evolved decks are also tested against the community decks. Overall trends in the winrates are explored and whether they hint at the transitivity of the deck space.

Because the decks are evolved and tested with opponent agents playing an aggro strategy, it is unclear whether the trends observed in the first two experiments will change when agents play with a different strategy. In section 8 the evolved decks play against each of the nine heroes who are playing starter decks for their class. While the success of a strategy is likely closely coupled with the deck, the goal is to see whether general trends persist.

Finally section 9 explores the degree to which strategies are transitive by evolving an additional set of decks where the opponents are playing an aggro strategy, but their decks are those originally evolved in section 6. If it would be impossible to evolve a deck that beats the evolved opponent deck more than 50% of the time, results would indicate that the strategy space has a very high degree of transitivity, implying the possibility of finding a globally optimal strategy. On the other hand, if it would be possible to evolve a deck that wins 100% of the time of the target evolved deck a low degree of transitivity is implied, i.e. that one deck's performance against another says little about its performance against a third deck.

In all deck comparisons significance is calculated based on a binomial proportion confidence interval with a general tz-test [3] and an exact Clopper-Pearson confidence interval for the observed winrates. The winrates are tested against a 50% winrate null hypothesis over 10,000 samples. Through this calculation a percentage difference larger than one is significant. All of the experiments are run on a high performance cluster with 400 CPUs running in parallel.

6 EXPERIMENT ONE AND RESULTS: DECK BUILDING

While ultimately evolution may have the power to discover successful decks for high-level play, this experiment first answers the question of whether it can evolve decks for amateurs by restricting the set of cards to those available to players at the start of the game [2]. While players collect more powerful cards throughout the game, these cards are available to anyone who completes the tutorial. Similarly, rather than evolving decks for the nine possible heroes players can select, experiments address three of these heroes who are expected to perform well with an aggro strategy,

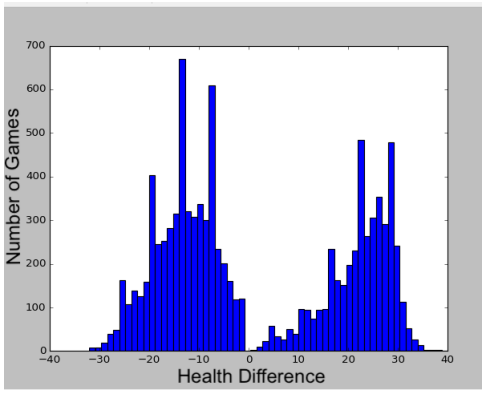


Figure 2: Health Distribution Between Two Community-Designed Decks The x-axis represents values of health differences, while the y-value represents the frequency with which games resulted in this health difference.

the hunter, paladin, and warlock. Future work will address deck building with a broader selection of heroes.

For each of these three heroes, the space of possible decks is explored through two runs of an evolution strategy (ES) algorithm ($\mu = 10$ and $\lambda = 100$) with a population size of 100 for each of the 100 generations. There is a 10% rate of elitism (i.e. 10% of the highest performing decks are copied into the new population) and the remaining 90% of the population are mutated copies of the top 10%. Mutation is performed by swapping a random card in the deck with another random card from the basic set. At least one card in the deck is swapped, and any time a swap is performed there is a 50% chance of an additional swap. Each evolved deck is evaluated through simulated gameplay against a hero of the same class playing with the starter deck.

Significant to evaluating decks in an ES is a measure of their fitness or quality. While winrate is an obvious factor in determining the quality of any given Hearthstone deck, there are many ways of calculating fitness with health differences and winrates. An exploratory experiment of 10,000 games between two community-made decks is shown in figure 2, where health differences between heroes are plotted on the x-axis and the number of games is shown on the y-axis. Because there is a valley rather than a bell curve at the mean and some asymmetry, this slightly skewed bimodal distribution indicates that the simply averaging health difference across multiple games may not be sufficient to establish deck quality as some subcurves could be easily over or undervalued. For instance, a small curve at the right end of the graph can overpower a much larger subcurve near the middle. To smooth the subcurves near the extremes of the distribution while still respecting the magnitude of the difference, deck fitness in this paper is calculated as the sigmoid of the health difference between two heroes at the end of a game. Health difference between heroes rather than winrate helps the ES to distinguish between strong wins and close games. Deck fitness in these experiments is determined by playing an experimentally determined four hundred games against an opponent with a fixed deck and strategy described in Section 4.2. Final evaluations are calculated by then averaging the mean of the sigmoid of the health differences with that of the deck's parent, which reduces the influence of stochasticity and noise. Each of the deckbuilding runs lasted 100 generations and approximately 33 hours.

6.1 Results

As shown in figure 3, best and average fitness in each of the runs for the heroes increases over time, indicating that decks effective against a fixed opponent can be evolved. While the fitnesses for both runs of the hunter and paladin converge after or before 100 generations, the second run of the warlock evolution could potentially increase in fitness with more time. Generally though, 100 generations is enough to converge to a solution.

While there are approximately 6.12×10^{20} [1] possible decks for any hero, interestingly each run has a least one initial starting deck with a best-fitness and approximate winrate over 60%. The existence of these randomly generated high-fitness decks in the initial generation could be in part because there are enough good cards to make almost any deck play well, or potentially because there are large enough sets of functionally similar cards.

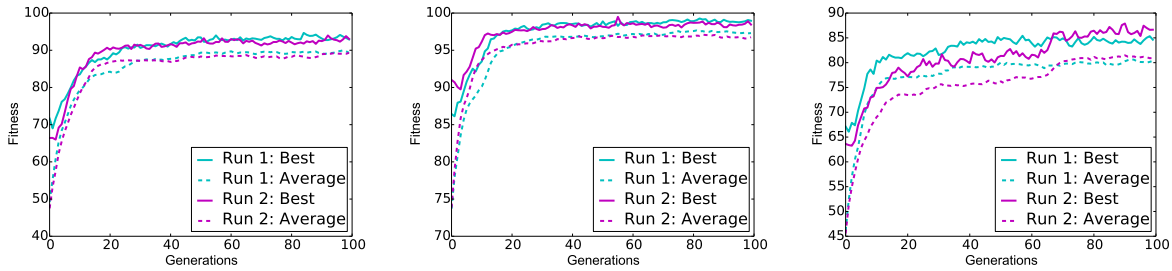
Because the fitness progression for the best-fitness decks is slightly jagged (i.e. it is possible for some earlier decks to outperform later decks), the highest fitness deck between each of the runs in any generation is selected for further exploration and testing (available at https://web.njit.edu/~ahoover/hearthstone/fdg2018/): generation 84 for the first run of the hunter (fitness = 94.67), generation 55 for the second run of the paladin (fitness = 99.49), and generation 90 of the first warlock run (fitness = 87.89).

Often decks can be partially assessed by their mana curve, which indicates the number of cards in the deck with particular mana costs. Because an aggro strategy prioritizes early-game wins and high cost mana cards are only available in late-game, successful aggro decks often have left-leaning manacurves. Figure 4 indicates of the evolved decks conform to the conventional wisdom of human players, emphasizing low mana costs for early play. For each of the three decks, over 86% of their cards cost four or fewer mana.

Available at https://web.njit.edu/~ahoover/hearthstone/fdg2018/, the three evolved decks share ten cards in common including one that is in each deck twice; Furthermore, most of these cards in common are considered good or viable for human-made aggro decks [1]. For instance, Sen'jin Shieldmasta is often considered a valuable defender against aggressive decks and playstyles, while the Raid Leader fills a unique niche among cards in the basic set. The Murloc Tidehunter, Razorfen Hunter, River Crocolisk are all considered reasonable additions to an aggressive deck of basic cards. In fact, only three are considered less desirable: Murloc Raider, Darkscale Healer, and Voodoo Doctor. While healing cards like the Voodoo Doctor and Darkscale healer naturally offer value to an aggressive playstyle as they simultaneously add attack power to the player's side while increasing their health, they are often absent from human-made aggro decks suggesting that they could be a valuable consideration when deckbuilding.

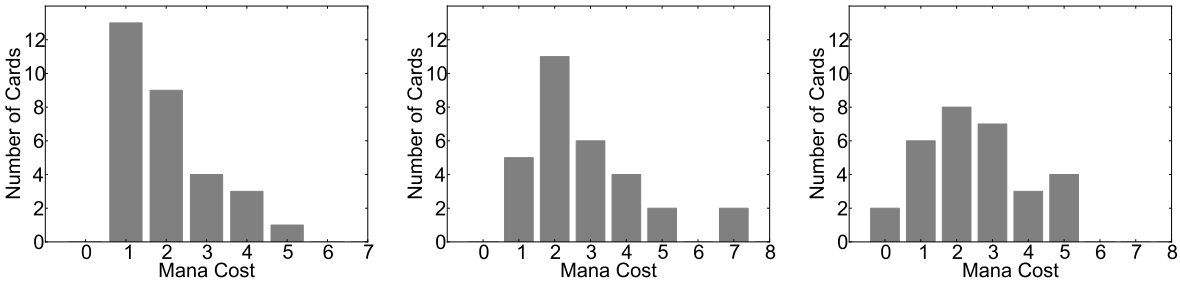
For at least one turn, taunt cards (e.g. Sen'jin Shieldmasta) help protect a player from an aggressive strategy as they must be eliminated before targeting its hero. The aggro strategy in figure 1

[1] Calculations performed by Matthew C. Fontaine with Python 3. The code and full explanation are available at https://web.njit.edu/ ahoover/hearthstone/fdg2018/



(a) Hunter Evo (b) Paladin Evo (c) Warlock Evo

Figure 3: *Fitness of Decks over 100 Generations.* Decks for each hero increase in the best and average fitnesses before converging to high fitness scores by 100 generations.



(a) Evolved Hunter (b) Evolved Paladin (c) Evolved Warlock

Figure 4: *Mana Curves of the Evolved Decks.* Conforming to conventional wisdom of successful aggro decks, the mana cost is low for most cards in each of the three evolved decks with over 86% of cards in each of the decks cost four or fewer mana.

	Taunt	Battlecry	Charge
Evo. Hunter	4	11	5
Evo. Paladin	6	12	0
Evo. Warlock	6	12	2

Table 1: *Special Abilities in the Evolved Decks.* While the aggro strategy favors taunt cards and these decks are close to maximizing those available in the basic set, there are a large number of battlecry cards and relatively few cards with charge.

encourages playing taunt minions, with four in the evolved hunter deck and six in the decks of the paladin and warlock (shown in table 1). While cards with battlecry are not accounted for in the aggro algorithm, there are 11 in the hunter deck, and 12 each in the decks of the paladin and warlock. Perhaps there are so many battlecry cards because even without explicitly promoting them, the turn-local agent exploits their immediate and beneficial affects. Generally minions must sleep the turn they are played (i.e. they cannot attack), yet battlecry minions provide immediate benefits because they impact the board state on the turn they are played.

Interestingly, while cards with charge also provide immediate benefits, relatively few are included in the evolved decks: five in the hunter deck, zero in the paladin deck, and two in the warlock deck. While there is one less charge card than taunt card available to all heroes in the basic set, perhaps fewer cards with charge are observed because they rarely last longer than a single turn and their value conventionally decreases dramatically when the opponent can block the charge with a taunt card (valued by both the opponent and player with an aggro strategy). Conceivably in future work, without the need for conventional wisdom or in-depth domain knowledge, computational exploration could help identify cards valuable to different playstyles.

7 EXPERIMENT TWO AND RESULTS: TESTING DECKS

To explore the specificity and quality of the generated decks, the evolved hunter, paladin, and warlock decks are each played against sets of decks designed for different purposes: the starter and community decks https://web.njit.edu/~ahoover/hearthstone/fdg2018/. The starter decks are composed of basic cards that are a straightforward collection of minions and spells, while the community decks are selected by advanced Hearthstone players and enthusiasts and considered to be more competitive than the starter decks that Blizzard initially introduces. Each evolved deck then plays an experimentally derived 20,000 games against each of the nine heroes playing the starter and community decks (10,00 games per deck).

The evolved decks are also played against each other to explore whether properties observed against the starter and community decks hold when played against each other.

7.1 Results

Table 2 shows results from the gameplay against all nine heroes, indicating that the evolved decks (i.e. the rows in table 2) significantly outplay the opponents playing the starter decks (i.e. the columns). Corresponding starter decks as players are also shown in the rows to illustrate the magnitude of difference in winrate between players with starter decks and those with evolved decks against all opponent heroes. Similarly, the evolved decks also outplay the community-crafted decks by larger margins than both the starter and community decks as shown in table 3.

From the distributions of winrates in tables 2,3, it appears that opponent difficulty is mostly invariant with respect to the player deck. That is, an opponent that is difficult for one deck seems to be difficult for all and vice versa. For example, the starter and evolved decks all perform best against the starter druid, hunter, mage, and priest and worst against the starter warlock and warrior. Similarly the starter, community, and evolved decks perform best against the community druid, hunter, and mage decks while struggling to win against the community paladin, shaman and warrior. That is decks that are difficult for one player, tend to be difficult for most.

Interestingly, after an additional set of 10,000 games played between each of the evolved decks shown in table 4, it is observed that the evolved deck's performance against starter and community deck wielding opponents correlates with the evolved decks performance against the other evolved decks. For example, the evolved paladin deck has some of the highest average winrates against the starter decks in table 2 (i.e. its winrates exceed most of the other evolved decks) where the evolved warlock is close but not quite as high. The evolved decks played against the community decks in table 3 show similar trends, where the evolved paladin deck has the highest winrates, with the evolved warlock coming in at a close second, and the evolved hunter last.

Similarly, the evolved paladin deck has the highest average winrate among evolved decks in table 4, where it more easily beats the evolved hunter deck than the evolved warlock. Likewise the warlock does its best against the hunter, once again suggesting the possibility that some universally optimal deck exists that can win against all other decks in the space of the basic set of cards. While the experiments are specific to the basic card set and the aggro strategy in SabberStone, results suggest that this approach could be expanded to make more generalizable conclusions by expanding the set of cards and strategies evaluated.

8 EXPERIMENT THREE AND RESULTS: EXPLORING THE IMPACT OF STRATEGY ON PERFORMANCE

Because the generated decks were evolved using only the aggro strategy, it is possible that evolution over-optimized decks to play with this specific heuristic. Although less valuable in high-level play, strategic universality is a desirable trait for beginner-friendly decks, allowing more amateur players to perform reasonably. To test for strategic universality of the evolved decks, a third experiment is proposed where the evolved decks are played against the starter decks in an additional set of 10,000 games. However rather than an aggro strategy, both players made decisions based on a basic strategy called control as represented by SabberStone's control heuristic.

8.1 Results

Similar to the results in table 4, table 5 illustrates that the paladin deck outperforms both the warlock and hunter decks where the evolved paladin's highest winrate is 89.40%, with the hunter's at 79.78% and the warlock's at 73.00%, all against the starter mage. These winrates suggest that the control heuristic has some degree of compatibility with the aggro heuristics, where the evolved paladin also outplayed the evolved hunter and evolved warlock against the starter deck experiments in table 2 and in the community decks in table 3. In these experiments, the evolved decks all did particularly well against the starter and community druid, hunter, and mage decks with similar winrates. Not only do these results point toward some amount of transitivity in the deck space, but that the trends in the results in section 7 are independent of these particular strategies.

9 EXPERIMENT FOUR AND RESULTS: CONTINUING EVOLUTION

With the same parameters as the experiment in section 6, this experiment further explores the transitivity of the deck space. However, rather than playing against the starter deck for the class, the population was evolved against the heroes playing highest performing decks of the previous evolution experiment, available at https://web.njit.edu/~ahoover/hearthstone/fdg2018/. Deck qualities are assessed through additional experiments against the starter and community decks.

9.1 Results

Interestingly, while the initial population for this second set of deck evolution was random, the new set of evolution experiments performed close to those originally evolved when played against the starter decks in section 7. Against both the starter and community decks (shown in tables 67), all of the evolved decks tended to perform best against the starter druid, hunter, and mage, with the community rogue also fairly easy to beat. This strong correlation between winrates against a variety of opponents indicates that winrate against one opponent is predictive of winrate against another, thereby providing more evidence to indicate the presence of transitivity and a universal optimum in the deck space of the basic cards in Hearthstone. While these findings apply specifically to the space of basic cards, future work will explore a broader selection of cards.

Results from playing all of the evolved decks together show the same trends observed when playing the originally evolved decks against each other (table 8), that is hunter decks were easiest to beat, followed by the warlock decks. The evolved paladin decks were the hardest. Excluding games played with identical heroes and decks, all decks play best against the hunter evolved against the previously evolved hunter deck, represented as evo. evo. hunter. They then performed second best against the originally evolved hunter (evo. hunter), evo. warlock, evo. evo. warlock, evo. paladin,

	Str. Druid	Str. Hunter	Str. Mage	Str. Paladin	Str. Priest	Str. Rogue	Str. Shaman	Str. Warlock	Str. Warrior
Str. Hunter	67.84%	N/A%	54.07%	43.77%	55.94%	42.47%	42.57%	22.54%	25.16%
Evo. Hunter	94.93%	92.45%	92.14%	88.65%	90.10%	86.41%	87.11%	72.37%	75.92%
Str. Paladin	61.79%	55.94%	65.61%	N/A%	56.06%	45.70%	51.44%	25.68%	30.20%
Evo. Paladin	98.03%	97.30%	97.97%	97.60%	96.30%	96.55%	97.07%	83.61%	89.36%
Str. Warlock	76.70%	77.06%	82.11%	74.63%	74.52%	74.01%	67.45%	N/A	52.94%
Evo. Warlock	95.71%	94.43%	95.65%	93.78%	93.45%	93.77%	91.46%	85.16%	80.62%

Table 2: *Evolved Decks vs. Starter Decks.* After 10,000 games between each evolved deck abbreviated Evo. Hunter, Evo. Paladin, and Evo. Warlock in the rows on the left and the Starter decks of the nine heroes abbreviated with Str. each of the nine columns, the evolved decks continually outperform the starter decks that they played during evolution. The performance of the original starter decks for each of the evolved decks is also shown for comparison.

	Co. Druid	Co. Hunter	Co. Mage	Co. Paladin	Co. Priest	Co. Rogue	Co. Shaman	Co. Warlock	Co. Warrior
Str. Hunter	70.79%	59.41%	71.26%	21.54%	60.98%	47.32%	23.59%	32.34%	22.68%
Co. Hunter	66.92%	50.17%	57.39%	15.03%	60.72%	37.41%	14.66%	26.85%	16.31%
Evo. Hunter	93.62%	95.89%	96.92%	68.08%	90.07%	90.59%	76.84%	78.05%	73.89%
Str. Paladin	66.04%	69.06%	82.17%	23.76%	55.39%	55.91%	25.84%	34.64%	25.4%
Co. Paladin	85.56%	85.89%	90.69%	50.28%	79.87%	80.64%	54.30%	62.99%	54.85%
Evo. Paladin	96.88%	98.66%	99.44%	76.61%	93.80%	96.42%	81.98%	84.39%	85.17%
Str. Warlock	79.76%	82.65%	91.03%	44.12%	77.39%	77.18%	44.86%	57.16%	47.05%
Co. Warlock	70.17%	72.46%	81.12%	36.67%	68.59%	67.91%	34.00%	50.47%	37.44%
Evo. Warlock	94.81%	97.18%	98.62%	68.51%	91.20%	93.14%	75.45%	83.78%	77.46%

Table 3: *Evolved Decks vs. Community Decks.* Fitnesses after 10,000 games played between the evolved decks (and corresponding starter and community decks), are shown in each of the rows corresponding to heroes playing the community-made decks. The community decks are player-designed sets of basic cards designed to perform better than those in the starter decks and are available at http://icy-veins.com**. Each row represents an evaluated deck, while each column corresponds to the class of the deck that is being played against.**

	Evo. Hunter	Evo. Paladin	Evo. Warlock
Evo. Hunter	49.74%	26.14%	42.18%
Evo. Paladin	72.85%	50.24%	66.86%
Evo. Warlock	57.52%	32.22%	51.50%

Table 4: *Evolved Decks vs. Evolved Decks.* After playing 10,000 games against each other, the paladin appears to be the strongest evolved deck, followed by the warlock, and hunter.

and finally the evo. evo. paladin. In only three cases did the evo. evo. deck perform significantly worse, and those were all with the evolved hunter decks against the three decks that were the hardest for all heroes to beat.

Given that the secondarily evolved decks performed better than evo. decks against almost all of the starter and community opponents (including the community decks absent from either evolutionary process), it appears important which deck the opponent is playing during evolution. While decks evolved against weaker opponents naturally win more easily against them, each of the evolved decks decrease in average fitness when playing the second set of evolved decks. Perhaps because the evo. paladin and warlock decks were stronger to begin with, the evo. evo. paladin and warlock decks were able to out perform the evo. and evo. evo. hunter decks on average. The evo. hunter averaged 39.35% against the evo. opponents, while the evo. warlock averaged 47.08%, and the evo. paladin 63.61%. While the evo. evo. hunter maintained that average across the evo. evo. opponents, both the evo. evo. paladin and the evo. evo. warlock increased their average fitness to 70.04 and 55.49. While the hero class may be an important issue with respect to these experiments, perhaps it was evolution against a much stronger opponent allowed the warlock and paladin experiments to generate stronger decks and a weak opponent force decks into a local optimum. Similarly, it is important to avoid starting with too strong of an opponent, as that will have the similar but opposite effect of pushing the candidates to the lower bound. Future work may benefit from periodically replacing the opponent deck during evolution to progressively increase the quality upper bound.

10 DISCUSSION AND FUTURE WORK

While limiting the cards available to players and restricting the gameplay strategies permits a basic exploration of the simplest space of Hearthstone decks, it is important to be careful when generalizing about the rest of the cards and strategies. Results point to the presence of transitivity and a universal optimum in the deck space of the *basic* cards in Hearthstone, but may not reflect properties of decks made of cards with more complicated abilities. Similarly, while the trends appear to hold regardless of the three heroes playing, it's conceivable that unique hero abilities and cards

	Str. Druid	Str. Hunter	Str. Mage	Str. Paladin	Str. Priest	Str. Rogue	Str. Shaman	Str. Warlock	Str. Warrior
Str. Hunter	50.08%	50.41%	78.88%	70.34%	66.47%	71.75%	78.90%	56.93%	38.23%
Co. Hunter	70.98%	59.34%	81.00%	73.25%	78.49%	70.69%	74.93%	59.56%	49.98%
Evo. Hunter	43.94%	64.12%	79.78%	50.59%	56.98%	56.77%	65.29%	50.19%	42.29%
Str. Paladin	30.80%	28.76%	61.10%	49.96%	56.12%	40.50%	68.01%	38.52%	27.72%
Co. Paladin	74.08%	63.48%	83.18%	78.19%	78.74%	66.98%	86.51%	67.88%	62.22%
Evo. Paladin	70.72%	81.63%	89.40%	72.58%	77.61%	85.09%	75.83%	62.25%	70.66%
Str. Warlock	47.58%	43.66%	63.54%	62.16%	68.87%	62.05%	76.27%	50.76%	43.24%
Co. Warlock	58.57%	54.19%	70.37%	76.64%	75.38%	68.27%	80.64%	60.63%	61.51%
Evo Warlock	48.38%	60.11%	73.00%	48.29%	62.67%	65.55%	54.64%	59.53%	43.32%

Table 5: *Exploring Effect of Strategy on Performance (Evolved Decks vs. Starter Decks).* The results of playing the decks against the starter set of decks with both players using a control heuristic rather than an aggro one.

	Str. Druid	Str. Hunter	Str. Mage	Str. Paladin	Str. Priest	Str. Rogue	Str. Shaman	Str. Warlock	Str. Warrior
Evo. Hunter	94.93%	92.45%	92.14%	88.65%	90.10%	86.41%	87.11%	72.37%	75.92%
Evo. Evo. Hunter	93.39%	92.69%	92.58%	89.13%	89.42%	87.96%	88.11%	72.13%	70.92%
Evo. Paladin	98.03%	97.30%	97.97%	97.60%	96.30%	96.55%	97.07%	83.61%	89.36%
Evo. Evo. Paladin	98.76%	97.58%	98.39%	97.92%	97.10%	97.37%	97.52%	88.48%	92.19%
Evo. Warlock	95.71%	94.43%	95.65%	93.78%	93.45%	93.77%	91.46%	85.16%	80.62%
Evo. Evo. Warlock	97.24%	96.03%	96.88%	96.09%	94.74%	95.44%	94.67%	81.99%	86.91%

Table 6: *Evolved Decks vs. Starter.* Interestingly, the evo. evo. decks had similar winrates to the evo. decks against the starter decks even though the evo. evo. decks were evolved only indirectly against them. Evo. evo. decks played against heroes playing the evo. decks in the second set of experiments, yet maintained their advantages over the starter decks.

	Co. Druid	Co. Hunter	Co. Mage	Co. Paladin	Co. Priest	Co. Rogue	Co. Shaman	Co. Warlock	Co. Warrior
Evo. Hunter	93.62%	95.89%	96.92%	68.08%	90.07%	90.59%	76.84%	78.05%	73.89%
Evo. Evo. Hunter	90.59%	96.69%	97.78%	66.45%	86.41%	90.08%	75.93%	77.28%	74.24%
Evo. Paladin	96.88%	98.66%	99.44%	76.61%	93.80%	96.42%	81.98%	84.39%	85.17%
Evo. Evo. Paladin	98.37%	98.98%	99.52%	81.73%	95.86%	98.06%	88.65%	90.60%	90.29%
Evo. Warlock	94.81%	97.18%	98.62%	68.51%	91.20%	93.14%	75.45%	83.78%	77.46%
Evo. Evo. Warlock	96.73%	97.73%	99.03%	72.74%	93.35%	95.03%	76.70%	83.07%	81.21%

Table 7: *Evolved from Evolved Decks vs. Community.* After 10,000 games the evo. evo. decks have similar winrates against the community decks as the originally evolved evo. decks hinting toward the transitivity of space of decks.

	Evo. Hunter	Evo. Evo. Hunter	Evo. Paladin	Evo. Evo. Paladin	Evo. Warlock	Evo. Evo. Warlock
Evo. Hunter	49.74%	45.83%	26.14%	25.43%	42.18%	34.53%
Evo. Evo. Hunter	55.11%	50.60%	22.67%	22.20%	41.37%	31.29%
Evo. Paladin	72.85%	77.11%	50.24%	36.58%	66.86%	56.52%
Evo. Evo. Paladin	73.97%	77.76%	61.79%	49.64%	74.35%	69.17%
Evo. Warlock	57.52%	60.21%	32.22%	24.59%	51.50%	42.82%
Evo. Evo. Warlock	64.98%	68.38%	43.71%	30.68%	57.79%	49.84%

Table 8: *Evolved Decks vs. Evolved Decks.* Like previous results, the paladin decks perform better than the warlock and hunter decks. While the evo. evo. decks for the paladin and hunter play better than their corresponding evo. decks, the evo. evo. hunter is the easiest deck for all of the others to beat.

of the other six heroes could skew the trends. Future work will extend the approach to consider more cards, heroes, and strategies.

Although class cards can help balance gameplay between different heroes, each hero also has a unique ability. For instance, regardless of cards currently in the players' hands, at each turn the hunter can do two damage to the enemy hero, the paladin can summon a minion with one attack and one health, and the warlock can lose two health to gain a card. Given that the paladin decks contained one to two class cards and the hunter and warlock decks four each, it raises the question of how much the hero power impacts

success with different strategies. Even with few class cards, the paladin had higher average winrates than the others. Future work will explore the impact of hero abilities on deckbuilding and whether some heroes naturally excel when playing different strategies.

While SabberStone's aggro strategy is suspected to resemble that of humans, it is conceivable that it simply reflects the strategy of the programmer who created it. Further testing with human players could help confirm. Furthermore through a machine learning analysis of how humans play, strategies beyond those currently available in SabberStone can potentially be developed and tested. Finally conceivably without the need for conventional wisdom or in-depth domain knowledge, computational exploration could help identify cards valuable to different playstyles and help players identify those that may perform best with their own.

11 CONCLUSION

This paper presents a series of experiments designed to explore the design space of decks in Hearthstone. After building decks for the hunter, paladin, and warlock with an ES, these decks were then tested with an aggro play strategy against both starter decks (i.e. the decks played during the tutorial) and some composed by fans. Results illustrated that all of the evolved decks tended to perform best against the same opponent decks, pointing toward an element of transitivity in the deck space (i.e. that some decks are inherently better than others). Experiments with a control play strategy confirmed initial observations that some decks are more beatable than others, but with winrates lower than those observed in games played with aggro playstyle, indicating that different types of decks may be found when playing with different strategies. A final set of experiments evolved another set of decks, but the opponent was playing with a deck from the previous ES experiment. Almost all of the secondarily evolved decks had higher winrates with the exception of the hunter. When compared against each other, the paladin decks proved hardest to beat, followed by the warlock, and hunter. The perhaps most obvious result is that the composition of the deck has a major influence on winrate against standard decks.

ACKNOWLEDGMENTS

Special thanks to the Academic and Research Computing Systems at NJIT (including Glenn Wolosh, David Perel) and New York University's High Performance Cluster team. Also, thank you to SabberStone developers Jean Seong Bjorn Choe (Milva) and Cedric Decoster (darkfriend77) for their assistance with the SabberStone simulations and AI. Thank you to Matthew C. Fontaine who helped calculate the total number of possible decks per hero. The authors also acknowledge the support of CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil.

REFERENCES

[1] 2014. Icy Veins: News and Guides. (2014). https://www.icy-veins.com/
[2] 2018 (accessed April 28, 2018). Card Set Basic. (2018 (accessed April 28, 2018)). https://hearthstone.gamepedia.com/Card_set#Basic
[3] Douglas G Altman. 1990. *Practical Statistics for Medical Research.* CRC press.
[4] Sverre Johann Bjørke and Knut Aron Fludal. 2017. *Deckbuilding in Magic: The Gathering Using a Genetic Algorithm.* Master's thesis. Norwegian University of Science and Technology (NTNU).
[5] Fernando de Mesentier Silva, Scott Lee, Julian Togelius, and Andy Nealen. 2017. AI-based Playtesting of Contemporary Board Games. In *Proceedings of the 12th International Conference on the Foundations of Digital Games.* ACM, 13.
[6] Joris Dormans. 2011. Simulating Mechanics to Study Emergence in Games. *Artificial Intelligence in the Game Design Process* 2, 6.2 (2011), 5–2.
[7] Blizzard Entertainment. 2014. Hearthstone. Online. (2014).
[8] Cedric Decoster et. al. 2018 (accessed March 30, 2018). *Sabberstone.* https://github.com/HearthSim/SabberStone
[9] Pablo García-Sánchez, Alberto Tonda, Giovanni Squillero, Antonio Mora, and Juan J Merelo. 2016. Evolutionary Deckbuilding in Hearthstone. In *Computational Intelligence and Games (CIG), 2016 IEEE Conference on.* IEEE, 1–8.
[10] Vincent Hom and Joe Marks. 2007. Automatic Design of Balanced Board Games. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE).* 25–30.
[11] Britton Horn, Amy K Hoover, Yetunde Folajimi, Jackie Barnes, Casper Harteveld, and Gillian Smith. 2017. AI-assisted Analysis of Player Strategy across Level Progressions in a Puzzle Game. In *Proceedings of the Foundations of Digital Games.* ACM, 9.
[12] Aaron Isaksen, Dan Gopstein, Julian Togelius, and Andy Nealen. 2015. Discovering Unique Game Variants. In *Computational Creativity and Games Workshop at the 2015 International Conference on Computational Creativity.*
[13] Alexander Jaffe, Alex Miller, Erik Andersen, Yun-En Liu, Anna Karlin, and Zoran Popovic. 2012. Evaluating Competitive Game Balance with Restricted Play. In *Proceedings of the Eighth Artificial Intelligence and Interactive Digital Entertainment International Conference (AIIDE 2012).*
[14] Jan Jakubik. 2017. Evaluation of Hearthstone Game States with Neural Networks and Sparse Autoencoding. In *Computer Science and Information Systems (FedCSIS), 2017 Federated Conference on.* IEEE, 135–138.
[15] Andrzej Janusz, Tomasz Tajmajer, and Maciej Świechowski. 2017. Helping AI to Play Hearthstone: AAIA'17 Data Mining Challenge. In *Computer Science and Information Systems (FedCSIS), 2017 Federated Conference on.* IEEE, 121–125.
[16] Jonas Krucher. 2015. Algorithmically Balancing a Collectible Card Game. *Bachelor's Thesis. ETH Zurich* (2015).
[17] Antonios Liapis, Christoffer Holmgård, Georgios N Yannakakis, and Julian Togelius. 2015. Procedural Personas as Critics for Dungeon Generation. In *European Conference on the Applications of Evolutionary Computation.* Springer, 331–343.
[18] Tobias Mahlmann, Julian Togelius, and Georgios N Yannakakis. 2012. Evolving Card Sets towards Balancing Dominion. In *2012 IEEE Congress on Evolutionary Computation.* IEEE, 1–8.
[19] Thorbjørn S Nielsen, Gabriella AB Barros, Julian Togelius, and Mark J Nelson. 2015. General Video Game Evaluation using Relative Algorithm Performance Profiles. In *European Conference on the Applications of Evolutionary Computation.* Springer, 369–380.
[20] André Santos, Pedro A Santos, and Francisco S Melo. 2017. Monte Carlo Tree Search Experiments in Hearthstone. In *Computational Intelligence and Games (CIG), 2017 IEEE Conference on.* IEEE, 272–279.
[21] Hans-Paul Schwefel and Günter Rudolph. 1995. Contemporary Evolution Strategies. In *European conference on artificial life.* Springer, 891–907.
[22] Andreas Stiegler, Keshav Dahal, Johannes Maucher, and Daniel Livingstone. 2017. Symbolic Reasoning for Hearthstone. *IEEE Transactions on Computational Intelligence and AI in Games* (2017).
[23] Julian Togelius, Georgios N Yannakakis, Kenneth O Stanley, and Cameron Browne. 2011. Search-based Procedural Content Generation: A Taxonomy and Survey. *IEEE Transactions on Computational Intelligence and AI in Games* 3, 3 (2011), 172–186.
[24] Vanessa Volz, Günter Rudolph, and Boris Naujoks. 2016. Demonstrating the Feasibility of Automatic Game Balancing. In *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference.* ACM, 269–276.
[25] Shuyi Zhang and Michael Buro. 2017. Improving Hearthstone AI by Learning High-level Rollout Policies and Bucketing Chance Node Events. In *Computational Intelligence and Games (CIG), 2017 IEEE Conference on.* IEEE, 309–316.
[26] Alexander Zook, Brent Harrison, and Mark O Riedl. 2015. Monte-Carlo Tree Search for Simulation-based Strategy Analysis. In *Proceedings of the 10th Conference on the Foundations of Digital Games.*