

Distinguishing Luck from Skill through Statistical Simulation: A Case Study

Zhaoyu Guo*, Irwin Khuu*, Kevin Zhu*, Jeffrey S. Rosenthal†, Frederic P. Schoenberg‡

(Version of April 14, 2019.)

Abstract

To investigate the perennial question of how to measure luck versus skill, we perform a detailed simulation study of Texas Hold'em poker hands. We define luck and skill as the player equity changes from dealing cards and from player betting decisions, respectively. We find that a careful definition of player equity leads to measurements of luck and skill which satisfy the statistical properties that we should expect them to. We conclude that our definitions of luck versus skill appear to be valid and accurate in the context of poker hands, and perhaps beyond to issues of luck and skill in other aspects of modern society.

1 Introduction

A deep and fundamental issue in our society is distinguishing the extent to which good and bad outcomes and achievements are the result of *luck* versus *skill*. Roughly speaking, luck refers to those influences which are beyond our control, i.e. which happen to us regardless of our efforts to affect them. By contrast, skill refers to those influences which are the direct result of our own conscious and deliberate actions as a result of our knowledge and experience and abilities. But how can we distinguish between these two forces in practice? If a computer programmer starts a dot-com company and gets rich, are the profits due to their own entrepreneurial skills, or to the luck of the market? If a rock group reaches number one on the charts, is that due to their musical talents, or to their good fortune in happening to connect with fickle and ever-changing entertainment trends?

The question of luck versus skill has been considered in many different contexts [8, 9, 21, 14], including many specifically about the game of *poker* (in which, among other things, distinguishing luck versus skill has implications for how profits are *taxed*) [6, 7, 19, 12, 11, 17].

*Equal contribution. University of Toronto. Emails: {kenny,khuu,kzhu}@utstat.toronto.edu

†Corresponding author. University of Toronto. Email: jeff@math.toronto.edu

‡University of California at Los Angeles. Email: frederic@stat.ucla.edu

Despite these many studies, it seems fair to say that luck and skill remain difficult quantities to measure or even to define.

In this paper, we investigate luck versus skill in Texas Hold'em poker games. We follow the general approach of [17] (see also [16]) that luck is what arises through the deal of the cards, while skill is what arises through players' decisions to bet or raise or fold. Furthermore, at any given stage of a poker hand, each player has an *equity* corresponding to their expected net profit on the hand conditional on the current hand situation. Increases in equity due to card deals count as luck, while increases in equity due to player decisions count as skill. (Of course, precise definitions of equity are challenging, and are discussed below.) The question then becomes, do these definitions give us a reasonable definition of luck versus skill, that satisfies the appropriate statistical properties that we would expect them to?

The article [17] expressed concern that such definitions of luck and skill would suffer from many drawbacks and limitations, including incorrect attributions of luck to skill, excessive correlations between the measured luck and skill, and challenges of appropriately defining a player's equity in the middle of a hand. However, in this paper, we demonstrate through extensive simulation studies that luck and skill can indeed be defined in ways which satisfy the expected statistical properties and appear to be useful and valid definitions of widespread applicability.

This paper is organised as follows. Section 2 provides needed background in the game of Texas Hold'em, Section 3 gives our precise definitions of how to measure luck and skill in this game, Section 4 describes our simulation set-up, Section 5 gives the results of our simulations, and Section 6 discusses the simulation results in detail. Section 7 closes with some conclusions and comments, including the observation that similar definitions of luck and skill may be applicable in far wider contexts in our society.

2 Texas Hold'em

Texas Hold'em is an example of an imperfect information game, in which players do not have knowledge of the entire game state and must conduct their actions through predictive analysis. However, while the game is imperfect, it is also a complete information game, so that all possible events in the game are known and can be planned for. Players have the agency to make decisions that will affect the final outcome of the game, allowing for a potential "skill" component, but do not have the ability to control the hidden information—which can affect their chance of winning the game—that is revealed as the game progresses, hence a "luck" component. For this paper, we will only be discussing the case of heads-up play, in which only two players are involved.

The goal of a hand of Texas Hold'em is to have the strongest hand by combining the player's own hole cards with the 5 cards on the board (assuming no one folds). The player with the winning hand claims all the chips in the pot (the total number of chips bet in the hand). If a player folds, the remaining player in the game claims all the chips in the pot. If the two players' hands are equal in strength upon revealing them, the pot is split equally between them. The game begins with the player in the small blind position putting in a

number of chips into the pot (called the *small blind*) and the player in the big blind puts in a larger amount, usually double that of the small blind (called the *big blind*). We will not be investigating antes in this paper. The dealer deals 2 hole cards to each of the players face down. The players can look at their own cards and take turns deciding whether to call (match the largest bet in the round), raise (place a bet larger than the largest bet in the round), fold (cede all the chips bet in the hand), or check (maintain the current bet amount, if that amount is equal to the highest amount on the board) starting with the player in the small blind. The players continue in sequence until all players have had at least one opportunity to raise and each player has bet the same amount. This round of betting is called the *pre-flop*. After the pre-flop, the dealer deals out 3 cards community cards face up. This is called the *flop*. Another round of betting begins as players decide whether to call, raise, fold or check. For this and the remaining rounds, the player in the big blind position bets first. We can see that the small blind has the position advantage of going later in the post-flop rounds. In each of the subsequent two rounds (called the *river* and the *turn*), the dealer deals out another community card which is followed by another round of betting.

Thus, we can see the two main components to acquire chips: making smart bets that either force the opposing player to fold or cause him to place bets on a weaker hand; and having the cards dealt such that the player's hand outperforms her opponent's hand. It is possible that a "bad" player makes "bad" move on a certain betting round but gets extremely lucky on the community cards and wins the pot. Hence the importance of separating skill and luck, to determine if the player is actually good or bad. The main difficulty in evaluating skill lies in valuing an action by a player. Since the players don't have perfect information, it is difficult for them to have an "optimal" strategy. For this paper, we will rely on the concepts of *equity* and *expected profit* in evaluating the players actions; these definitions will be further explained in the next section. By simulating individual hands from head-to-head matches of player routines from the R package 'holdem' [18], as well as strategies that we have developed ourselves [2], the effectiveness of these definitions will be evaluated.

Analysis in poker games are also found in [8], [6], and [21]. Levitt *et al.* proposed a data-driven approach [11, 12], while other approaches also include trying to estimate a Nash Equilibrium for the game [10]. An interesting paper by Van Loon *et al.* also discusses the game of online poker[19].

3 Definitions of Luck and Skill

The definitions of luck and skill provided in [17] are dependent on the poker concepts of *equity* and *expected profit*. Equity is the expected proportion of the pot a player earns under the assumptions of no future betting and using only the knowledge of cards in the players' hand and revealed cards on the board. It is calculated by multiplying their win probability with the size of the pot. The expected profit of the player is their equity in the hand subtracted by the cost of their betting actions.

The definitions of luck and skill are thus:

- Luck: the change in expected profit as cards are dealt in terms of chips.
- Skill: the change in expected profit during the betting rounds (i.e. as a result of betting actions) in terms of chips.

Borm and Van der Genugten [9] provide some reasonable criteria for the ideas discerning a game of luck from a game of skill, which we would like to verify with our definitions:

- The skill of a player should be measured as his average game result in the long run, i.e., in terms of expected result.
- For a game of skill it is necessary that these expected results vary among players.
- The fact that there is a difference between players with respect to their expected payoffs, does not immediately imply that the underlying game is a game of skill. For a game of skill it is sufficient that the chance elements involved do not prohibit these differences to be substantial.

Furthermore, the article [17] provides some possible concerns with the current definitions. We seek to evaluate these limitations and determine if they are significant enough to indicate alternative definitions should be considered.

- Situations can occur where a terrible player may gain expected profit during betting rounds against a better player, and attributing such gains to skill may be objectionable.
- Luck and skill can be correlated in practice. This would suggest that the given definitions form a poor demarcation between the two ideas.
- In simple equity calculations the assumption of no future betting or folding can lead to significant overestimates or underestimates of the equity gained/lost from winning/losing the hand.

The order of betting actions in a head-to-head match of Texas Hold'em is as follows: in the pre-flop phase, the person to the left of the big blind will make their bet, i.e. the small blind player. From the flop onwards, the big blind player will conduct their betting actions first.

Luck and skill gains and losses are evaluated on a per round basis. Thus a player may have individual rounds that result in skill losses as well as rounds that result in skill gains. This similarly applies to gains and losses due to luck (the deal, the flop, the turn and the river), and up to four components of the gains/losses due to skill (pre-flop betting, post-flop betting, post-turn betting, and post-river betting). Furthermore, the sum of the skill and luck gains of the winning player necessarily matches sum of the skill and luck losses of the other player and will equal the change in chip count at the end of the hand.

In the case of pre-flop expected profit calculations, the skill and luck gains are less obvious because the big blind and the small blind offer a different number of chips. In the case where

the small blind folds without raising or calling in the pre-flop, which would be a case in which the two players have not put in the same number of chips, the pre-flop equity is calculated to be the expected profit under the assumption that either the big blind and small blind both call, or the equity the small blind would have by folding, whichever is greater. For example, consider a head-up game with small blind m and big blind $2m$. Then the pre-flop equity of the big blind would be $4mp - 2m$. That is, the small blind calls the bet so the pot is increased to $4m$ and the big blind's equity would be the pot $4m$ multiplied by his win probability p , but $2m$ of the pot was added by the big blind and thus is subtracted when calculating expected profit. Meanwhile, the pre-flop equity for the small blind would be calculated by $\max\{4mq - 2m, -m\}$, where $q = 1 - p$ is the small blind's probability of winning the pot. The probability of ties are divided equally between the two players when calculating p and q . This definition defines increases in the size of the pot relative to the big blind, so that increasing the pot size by calling the blinds is not counted as skill.

A consequence of this clarification is that a player's gains and losses due to luck and skill will not necessarily match their opponents' change in chips due to luck and skill. Consider a case with blinds $m/2m$ in which the big blind has a pre-flop win probability of $p > 0.75$. His expected profit due to luck will be $4mp - 2m$ compared to the small blind's expected profit of $-m$, but $4mp - 2m > |-m|$, so the big blind's expected profit due to luck will be greater than the luck-based expected loss of the small blind if the small blind folds immediately. Furthermore, since the sum of the expected profits due to luck and skill equal the change in chips, the "excess" luck profit gained by the big blind will be balanced out by a negative expected profit due to skill, despite no action being taken by the big blind! As a result, over the course of many hands, the big blind's average expected profit due to skill will always be slightly higher, and average expected profit due to luck slightly lower, than those of the small blind. While this peculiarity is counterintuitive, its effect is minor enough to be nearly negligible due to the rarity of the circumstances and the relatively small size of the pot in such circumstances.

In any case, if the small blind does not immediately fold in the pre-flop, any future betting actions that cause the opposing player to fold are considered entirely as gains from skill since the win probability increases to 100% as a result of the winning player's betting action.

4 Simulation Setup

To investigate luck versus skill in poker, we considered using data from actual championship poker matches (cf. [11]), but found this to be problematic due to constantly changing circumstances in terms of chip counts, multiple players, etc. We also considered using data from Artificial Intelligence poker systems [1, 13, 5, 15, 20, 4, 10], but found this difficult due to such factors as widely varying abilities of the human opponents over small numbers of hands in the available data trials.

So, instead, we ran our own simulation studies, using some of the poker-playing routines in the R package 'holdem' [18] together with some additional poker-playing routines that we

wrote ourselves to illustrate different phenomena. This had the advantage that we could run simulated poker hands as often as desired, with complete control over the game conditions, thus leading to more accurate and insightful results.

We simulated hands of Texas Hold'em using six unique player strategies meant to perform different archetypes and play styles with different degrees of success. The six player algorithms include four newly-defined routines called GoodTight, GoodLoose, BadTight, and BadLoose, as well as the routines Tommy and Zelda from the 'holdem' package; the specifics of their strategies are described in Appendix A.

The conditions under which we conducted our simulations are as follows:

- The chips available to each player at each round will be 20,000.
- The blinds will be 50/100.
- The chip counts will be reset after every match so that betting strategies remain consistent.

In a head-to-head match, calculating the exact win probability of the players hands would involve the calculation of $\binom{48}{5} = \frac{48!}{5!(48-5)!} = 1.71$ million different combinations. In this case, we instead use Monte Carlo simulations with a sample size of $M = 2,000$ to obtain an estimate of the win probability. After the flop is revealed, the number of possible combinations for the turn and river are $\binom{45}{2} = 990$ and $\binom{44}{1} = 44$, respectively, so the win probabilities for all possible combinations are computed exactly (as in e.g. [3]). Of course, once the river is revealed, the win probability will be either 0, 1, or $\frac{1}{2}$ in the case of a tie.

For each pair of strategies, we calculate the mean luck and skill gains or losses over 3000 simulated hands, with each player playing 1500 hands as big blind and 1500 hands as small blind in each match up.

5 Simulation Results

All the simulations are conducted with 3000 hands, where each hand uses 2000 Monte Carlo iterations to calculate the approximate win probability in the pre-flop stage. We split the results into two sections: one where four newly developed player strategies face off against each other; and a section in which two players from the R package `holdem` play against the new strategies.

New Players

The four newly developed player strategies are designed to emulate two distinct strategies, either "tight" or "loose", with either high or low competence. The strategy names are GoodLoose, GoodTight, BadLoose, and BadTight. A detailed description of their design can be seen in the Appendix. Table 1 describes the results of the 3000 hand simulations.

Big Blind	Small Blind		Luck Profit		Skill Profit		Chips Change	
			BB	SB	BB	SB	BB	SB
GoodLoose	GoodTight	Mean (SE)	25.44 (169.76)	-25.08 (169.75)	23.96 (177.94)	-24.33 (177.94)	49.40 (242.53)	-49.40 (242.53)
GoodTight	GoodLoose	Mean (SE)	-96.62 (124.18)	97.06 (124.18)	-520.18 (193.33)	519.74 (193.33)	-616.80 (237.81)	616.80 (237.81)
GoodLoose	BadLoose	Mean (SE)	-129.84 (181.02)	130.38 (181.02)	2115.71 (258.42)	-2116.25 (258.44)	1985.87 (326.80)	-1985.87 (326.80)
BadLoose	GoodLoose	Mean (SE)	21.43 (148.55)	-21.02 (148.55)	394.91 (133.94)	-395.31 (133.93)	416.33 (202.60)	-416.33 (202.60)
GoodLoose	BadTight	Mean (SE)	-20.03 (7.70)	20.59 (7.70)	-371.77 (77.35)	371.21 (77.35)	-391.80 (80.54)	391.80 (80.54)
BadTight	GoodLoose	Mean (SE)	-3.93 (3.19)	4.46 (3.20)	-244.41 (66.24)	243.88 (66.24)	-248.33 (67.76)	248.33 (67.76)
BadLoose	BadTight	Mean (SE)	105.38 (76.15)	-104.92 (76.15)	-141.45 (43.70)	140.98 (43.70)	-36.07 (84.39)	36.07 (84.39)
BadTight	BadLoose	Mean (SE)	25.36 (77.27)	-25.00 (77.26)	-97.90 (50.07)	97.53 (50.05)	-72.53 (88.60)	72.53 (88.60)
BadLoose	GoodTight	Mean (SE)	-103.77 (86.30)	104.30 (86.30)	-240.76 (101.42)	240.23 (101.42)	-344.53 (134.81)	344.53 (134.81)
GoodTight	BadLoose	Mean (SE)	78.20 (106.14)	-77.66 (106.15)	-2454.07 (231.55)	2453.53 (231.54)	-2375.87 (267.55)	2375.87 (267.55)
GoodTight	BadTight	Mean (SE)	9.44 (11.41)	-8.82 (11.41)	-232.51 (80.92)	231.88 (80.92)	-223.07 (82.06)	223.07 (82.06)
BadTight	GoodTight	Mean (SE)	-13.68 (5.57)	13.92 (5.57)	-241.52 (75.08)	241.28 (75.08)	-255.2 (77.53)	255.2 (77.53)

Table 1: From 3000 simulated hands, average profit due to luck and skill in head-to-head match-ups, with switching blinds, between four player strategies.

Table 2 extends these results by also including the resulting correlation between luck and skill to examine the strength and direction of their relationship, as well as the ratio of the average results in order to gauge the approximate magnitude of the effects luck and skill have on overall results.

Big Blind	Small Blind	Luck-Skill ρ	Skill/Luck Ratio	
			BB	SB
GoodLoose	GoodTight	-0.027	0.942	0.970
GoodTight	GoodLoose	0.078	5.384	5.355
GoodLoose	BadLoose	0.077	-16.295	-16.231
BadLoose	GoodLoose	0.026	18.432	18.806
GoodLoose	BadTight	0.372	18.557	18.027
BadTight	GoodLoose	0.457	62.259	54.712
BadLoose	BadTight	-0.088	-1.342	-1.344
BadTight	BadLoose	-0.081	-3.860	-3.901
BadLoose	GoodTight	0.025	2.320	2.303
GoodTight	BadLoose	0.136	-31.380	-31.593
GoodTight	BadTight	0.030	-24.623	-26.301
BadTight	GoodTight	0.411	17.654	17.331

Table 2: From Table 1 results, corresponding correlation (ρ) between luck and skill, as well as the ratio of skill to luck average profits.

Tommy vs. New players

The strategy "Tommy" is one of the built-in strategies available in the R `holdem` package. Table 3 describes results of the 3000 hand simulations, while Table 4 examines the correlation between luck and skill for each hand, as well as the ratio of the average luck and skill profits.

Big Blind	Small Blind		Luck Profit		Skill Profit		Chips Change	
			BB	SB	BB	SB	BB	SB
Tommy	GoodLoose	Mean (SE)	-0.84 (1.89)	1.36 (1.90)	66.30 (2.02)	-66.83 (2.00)	65.47 (1.94)	-65.47 (1.94)
GoodLoose	Tommy	Mean (SE)	-12.78 (12.80)	13.28 (12.79)	40.81 (1.21)	-41.32 (1.19)	28.03 (13.39)	-28.03 (13.39)
Tommy	BadLoose	Mean (SE)	46.55 (108.44)	-46.14 (108.44)	-325.61 (57.31)	325.20 (57.37)	-279.07 (123.50)	279.07 (123.50)
BadLoose	Tommy	Mean (SE)	-6.98 (107.28)	7.50 (107.28)	-332.29 (53.47)	331.77 (53.47)	-339.27 (116.72)	339.27 (116.72)
Tommy	GoodTight	Mean (SE)	1.28 (2.13)	-0.75 (2.12)	38.62 (1.79)	-39.15 (1.81)	39.90 (2.37)	-39.90 (2.37)
GoodTight	Tommy	Mean (SE)	-1.58 (0.84)	2.10 (0.81)	40.58 (1.09)	-41.10 (1.07)	39.00 (1.01)	-39.00 (1.01)
Tommy	BadTight	Mean (SE)	12.31 (7.80)	-11.87 (7.80)	14.83 (11.48)	-15.26 (11.48)	27.13 (19.02)	-27.13 (19.02)
BadTight	Tommy	Mean (SE)	20.43 (24.05)	-19.85 (24.04)	47.21 (8.19)	-47.78 (8.20)	67.63 (26.66)	-67.63 (26.66)

Table 3: From 3000 simulated hands, average profit due to luck and skill in head-to-head match-ups, with switching blinds, between "Tommy" strategy and four new player strategies.

Big Blind	Small Blind	Skill & Luck ρ		Skill/Luck Ratio	
		BB	SB	BB	SB
Tommy	GoodLoose	-0.509	-0.509	-79.285	-48.977
GoodLoose	Tommy	0.462	0.462	-3.11	-3.193
Tommy	BadLoose	0.016	0.016	-6.996	-7.049
BadLoose	Tommy	-0.065	-0.065	44.239	47.602
Tommy	GoodTight	-0.278	-0.278	30.113	51.871
GoodTight	Tommy	-0.461	-0.461	-19.577	-25.732
Tommy	BadTight	0.944	0.944	1.205	1.286
BadTight	Tommy	0.165	0.165	2.406	2.311

Table 4: From Table 3 results, corresponding correlation between luck and skill as well as skill to luck ratio of average profits.

Zelda vs. New players

The strategy "Zelda" is one of the built-in strategies available in the R `holdem` package. Table 5 describes results of the 3000 hand simulations, while Table 6 examines the correlation between luck and skill for each hand, as well as the ratio of the average luck and skill profits.

Big Blind	Small Blind		Luck Profit		Skill Profit		Chips Change	
			BB	SB	BB	SB	BB	SB
Zelda	GoodLoose	Mean (SE)	-37.98 (58.46)	38.35 (58.46)	-911.05 (112.54)	910.68 (112.54)	-949.03 (129.14)	949.03 (129.14)
GoodLoose	Zelda	Mean (SE)	5.93 (36.78)	-5.56 (36.78)	-111.05 (41.28)	110.69 (41.28)	-105.13 (51.72)	105.13 (51.72)
Zelda	BadLoose	Mean (SE)	14.99 (76.94)	-14.52 (76.94)	-514.76 (92.12)	514.28 (92.16)	-499.76 (121.08)	499.76 (121.08)
BadLoose	Zelda	Mean (SE)	116.63 (83.21)	-116.22 (83.22)	-556.47 (89.53)	556.06 (89.53)	-439.84 (115.99)	439.84 (115.99)
Zelda	GoodTight	Mean (SE)	22.98 (40.95)	-22.45 (40.95)	-657.80 (109.32)	657.28 (109.32)	-634.82 (123.40)	634.82 (123.40)
GoodTight	Zelda	Mean (SE)	-22.02 (23.81)	22.52 (23.81)	-3.22 (31.13)	2.71 (31.13)	-25.23 (40.04)	25.23 (40.04)
Zelda	BadTight	Mean (SE)	1.83 (5.28)	-1.49 (5.28)	-51.50 (48.46)	51.06 (48.46)	-49.67 (49.95)	49.67 (49.95)
BadTight	Zelda	Mean (SE)	-0.70 (1.08)	1.06 (1.07)	22.20 (23.02)	-22.56 (23.02)	21.5 (23.15)	-21.5 (23.15)

Table 5: From 3000 simulated hands, average profit due to luck and skill in head-to-head match-ups, with switching blinds, between "Zelda" strategy and four new player strategies.

Big Blind	Small Blind	Skill & Luck ρ		Skill/Luck Ratio	
		BB	SB	BB	SB
Zelda	GoodLoose	0.045	0.045	23.988	23.747
GoodLoose	Zelda	-0.126	-0.126	-18.735	-19.906
Zelda	BadLoose	0.170	0.170	-34.338	-35.426
BadLoose	Zelda	-0.100	-0.100	-4.771	-4.784
Zelda	GoodTight	0.179	0.179	0.146	0.121
GoodTight	Zelda	0.045	0.045	-28.628	-29.271
Zelda	BadTight	0.232	0.232	-28.092	-36.545
BadTight	Zelda	0.094	0.094	-31.865	-21.241

Table 6: From Table 5 results, corresponding correlation between luck and skill as well as skill to luck ratio of average profits.

6 Discussion of Simulation Results

Aside from two match ups, GoodLoose (BB) vs. GoodTight (SB) and BadLoose (BB) vs. BadTight (SB), the simulation results for the new players show a large ratio between skill profit and luck profit. This provides evidence that our definitions are capturing the skill and luck of the players. For perfect definitions, the mean of luck profit should go to zero

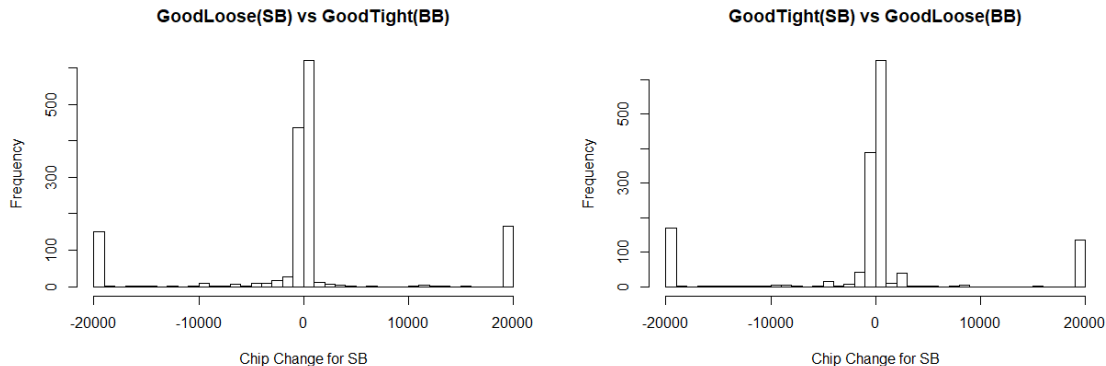
as the number of hands increases. The table shows almost a tie between the two players in the match-ups of GoodLoose vs GoodTight and BadLoose vs BadTight, which may explain the large skill luck ratio. Furthermore, we see that the standard error of the skill profit is less than that of the chip change, suggesting an improvement of measuring skill compared to only using the chip count. One of the issues with the current definition of skill is that it seems the skill profit is highly dependent on the opponent’s strategy and the positions, which makes the definition somewhat inconsistent.

Additionally, the magnitude of the correlation between luck and skill ρ is less than 0.1 in the majority of the match-ups. Furthermore, in those few cases when the correlation is large, the ratio of skill to luck is very high, indicating that the luck has almost entirely cancelled out anyway. This result indicates that the relationship between luck and skill suggested in [17] is not as significant an issue as previously assumed.

We next investigate three specific match-ups: GoodLoose vs. GoodTight, GoodLoose vs. BadTight, and GoodTight vs. BadTight. The first match-up helps illustrate our definitions in action, whereas the latter two seem to give counterintuitive results.

GoodLoose vs GoodTight:

The main contrast between playing styles among Poker players is how loose or tight (conservative) one plays. Of course, in reality, players play somewhere near the middle of this spectrum, and often switch up their strategies. In this paper, we will assume the players always play tight or loose and investigate how they do against each other. Details about the players’ strategies can be found in the Appendix. In a nutshell, compared to GoodTight, GoodLoose will bet slightly more (or call) when it has a hand or draw. GoodLoose will also have a higher chance of calling when it does not have a hand. GoodTight does not bet (when there is no bet to call) or raise if it does not have a hand, whereas GoodLoose will sometimes. We will start with investigating GoodTight vs GoodLoose since, as seen earlier, this match up gave a higher skill to luck ratio. Since the table only has the means of the resulting hands, we further break down the results to individual hands.



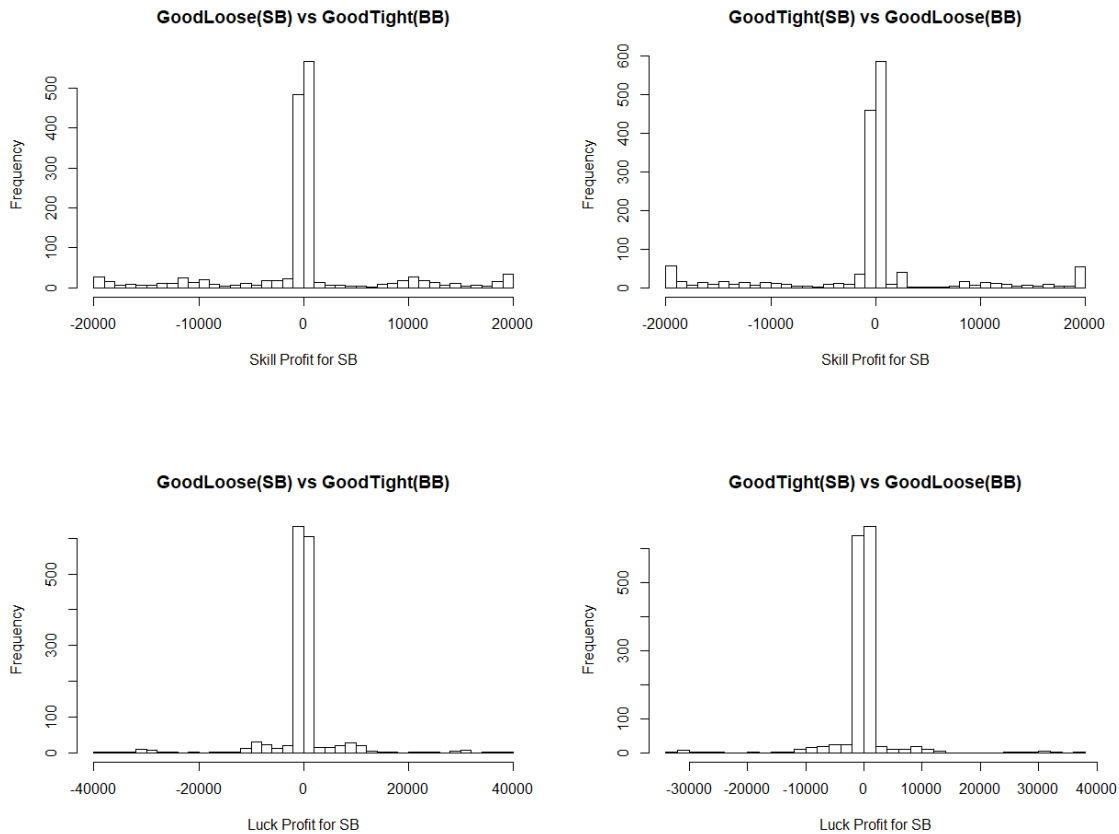


Figure 1: Individual hands breakdown of GoodLoose vs GoodTight (1500 hands each match up).

As seen from the histograms in Figure 1, the player in the small blind seem to win most of the small pots (winner claims big blind amount) but GoodLoose wins more of the all-in pots regardless of position. The luck profit seems to be pretty symmetric.

However, if we take a look at all the non all-in pots, we see that the player in the big blind actually wins more, as seen below in Table 7.

Small Blind		SB Luck	BB Luck	SB Skill	BB Skill	SB Chip	BB Chip
GoodLoose	Mean	-8.78	9.14	-182.16	181.81	-190.95	190.95
1183 hands	(SE)	(21.06)	(21.06)	(38.25)	(38.25)	(39.11)	(39.11)
GoodTight	Mean	-30.68	31.03	-158.09	157.74	-188.77	188.77
1193 hands	(SE)	(18.31)	(18.31)	(37.14)	(37.14)	(36.98)	(36.98)

Table 7: Skill and Luck Equity for GoodLoose vs GoodTight in hands where the pot is less than 10000.

When GoodLoose is in the big blinds and neither player has a good hand, there is a chance that GoodLoose bets without a hand (bluffs), inducing GoodTight to fold and thus winning the pot. Table 8 provides a detailed breakdown of how the hands play out.

	GoodLoose(SB)	GoodTight(SB)
SB folds after BB raises during PreFlop	0.026	0.018
BB folds after SB raises during PreFlop	0.017	0.014
BB folds during Flop	0.303	0.133
SB folds during Flop	0.189	0.300
BB folds during River	0.079	0.022
SB folds during River	0.019	0.088
BB folds during Turn	0.019	0.015
SB folds during Turn	0.019	0.033
Showdown	0.327	0.376

Table 8: Percentage of how games end among non all-in pots. The first column is when GoodLoose is in the Small Blind (1183 hands), and the second is when GoodTight is in the Small Blind (1193 hands).

On the other hand, when GoodLoose is in the small blinds, it is likely to call with a bad hand after GoodTight raised with a good hand, resulting in GoodTight winning more on average in the big blind position.

GoodLoose vs BadTight:

The BadLoose and BadTight players are not necessarily “bad”, but just play more extreme on the spectrum of loose vs tight. Essentially, BadLoose will call any bet and will always bet if the amount to call is zero. BadTight on the other hand will only raise all in if it has a really good hand and fold or check otherwise. It seems GoodLoose vs BadTight and GoodTight vs BadTight are scenarios where the “bad” or more extreme play styles beat the “good” player. I will first investigate GoodLoose vs BadTight.

There are only 3 resulting scenarios from this match up: (i) a player wins 100 (BB amount), (ii) The player in the BB (BadTight) wins the all-in pot, (iii) a tie. The positive skill profit (and chip profit) in the results table is mainly due to (ii). This is shown in Figure 2.

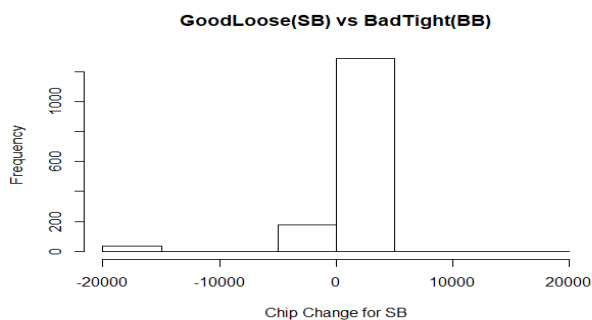


Figure 2: Individual hands breakdown of GoodLoose vs BadTight (1500 hands each match up).

If we remove the all-in pots, it shows GoodLoose as the better player.

Small Blind		SB Luck	BB Luck	SB Skill	BB Skill	SB Chip	BB Chip
GoodLoose	Mean	1.22	-0.65	75.54	-76.11	76.76	-76.76
1463 hands	(SE)	(1.33)	(1.32)	(1.32)	(1.32)	(1.16)	(1.16)

Table 9: Skill and Luck Equity for GoodLoose vs BadTight after removing all in pots.

GoodTight vs BadTight:

The outcome here is almost the same as the previous (GoodLoose vs BadTight), except the small blind (GoodTight) also wins all-in pots but much less than the big blind. This is illustrated in Figure 3.

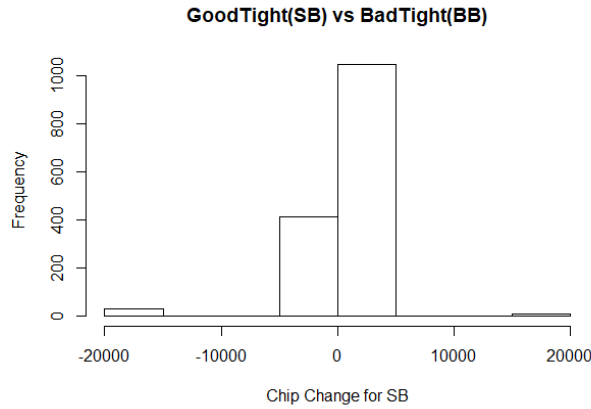


Figure 3: Individual hands breakdown of GoodTight vs BadTight (1500 hands each match up).

Table 10 shows that if we remove the all-in pots, it shows GoodTight as the better player.

Small Blind		SB Luck	BB Luck	SB Skill	BB Skill	SB Chip	BB Chip
GoodTight	Mean	17.09	-16.45	27.64	-28.28	44.73	-44.73
1462 hands	(SE)	(7.68)	(7.68)	(7.65)	(7.65)	(1.62)	(1.62)

Table 10: Skill and Luck Equity for GoodTight vs BadTight after removing all in pots.

Zelda vs New Players:

We play our new players against Zelda, a player from the *Hold'em* package. Zelda's strategy definition can be found in the Appendix. The match ups here are interesting, as

they all end up in a similar scenarios as for when BadTight played the other new players, as seen in Figure 4.

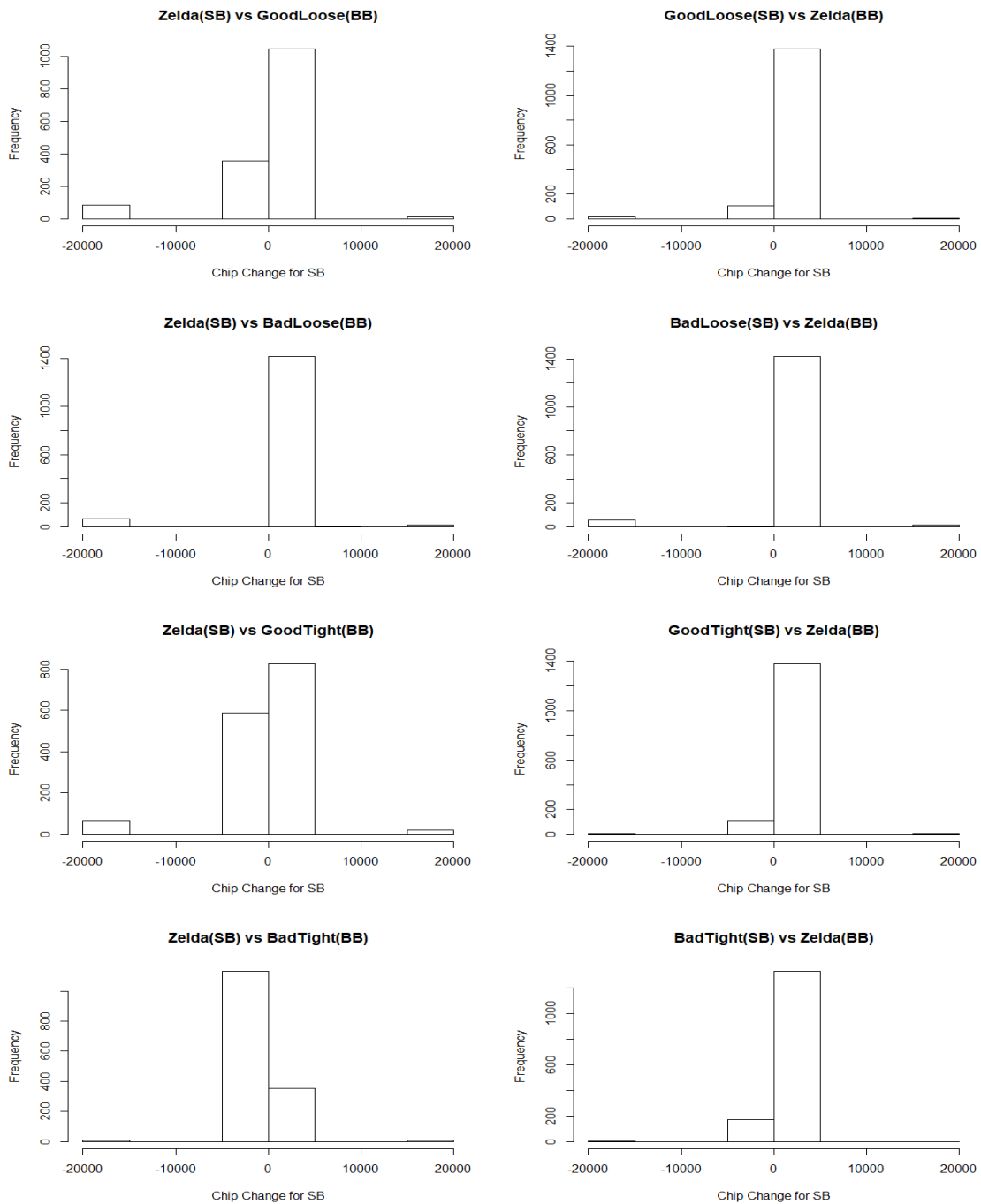


Figure 4: Individual hands breakdown of new players vs Zelda (1500 hands each match up).

In the case of Zelda vs BadLoose, other than some all-in pots, Zelda seems to never lose. These results are also due to Zelda having a high rate of directly folding during pre-flop from the small blind position. This is shown in Table 11.

Match Up (SB vs. BB)	Fold Rate
Zelda vs GoodLoose	0.843
Zelda vs BadLoose	0.853
Zelda vs GoodTight	0.866
Zelda vs BadTight	0.853

Table 11: Rate of directly folding from the small blind position pre-flop.

As an aside, we note that in our simulations, Zelda generally performs poorly at poker in the sense of losing many chips on average, even when playing against our new players who were designed solely to illustrate different behaviours. This illustrates the difficulty in designing simple poker-playing strategies like Zelda to be able to win at the subtle game of poker.

Tommy vs New Players:

Tommy is another player from the *Hold'em* package we used. Tommy's strategy is much more simple than Zelda's. Tommy goes all in whenever it has a pocket pair. All of the new players seem to be quite evenly matched with Tommy after 1500 hands, as the skill profit (or chip change) are lower than the Big Blind amount (100). The only exception is BadLoose, to where the player in big blind wins more. The individual hands breakdown are shown in Figure 5.

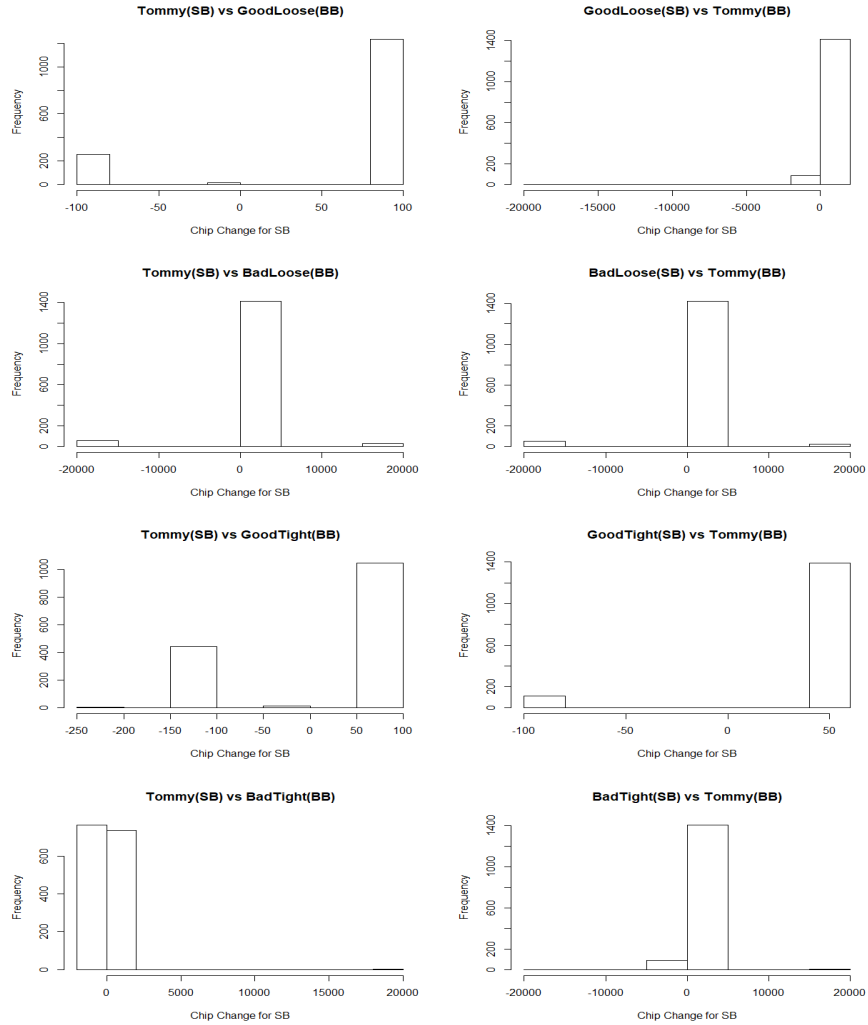


Figure 5: Individual hands breakdown of new players vs Tommy (1500 hands each match up).

Table 12 describes the rate at which the small blind folds in the pre-flop stage of the hand.

Match Up (SB vs. BB)	Fold Rate
Tommy vs GoodLoose	0.843
Tommy vs BadLoose	0.948
Tommy vs GoodTight	0.927
Tommy vs BadTight	0.938

Table 12: Rate of directly folding from the small blind position pre-flop.

Convergence Plots:

For a different perspective, we consider how the mean skill per hand, and the mean luck per hand, vary as the number of hands increase. For small numbers of hands, these quantities are highly variable due to the randomness of the simulations. However, for larger numbers of hands, the mean luck per hand should hopefully converge to zero, while the mean skill per hand should converge to a true measure of the relative skill levels of the two players.

We illustrate this for all of the matchups involving the BadLoose player. In our plots below, the x-axis is the number of hands played, and the y-axis is mean luck (red) and skill (green) gain per hand.

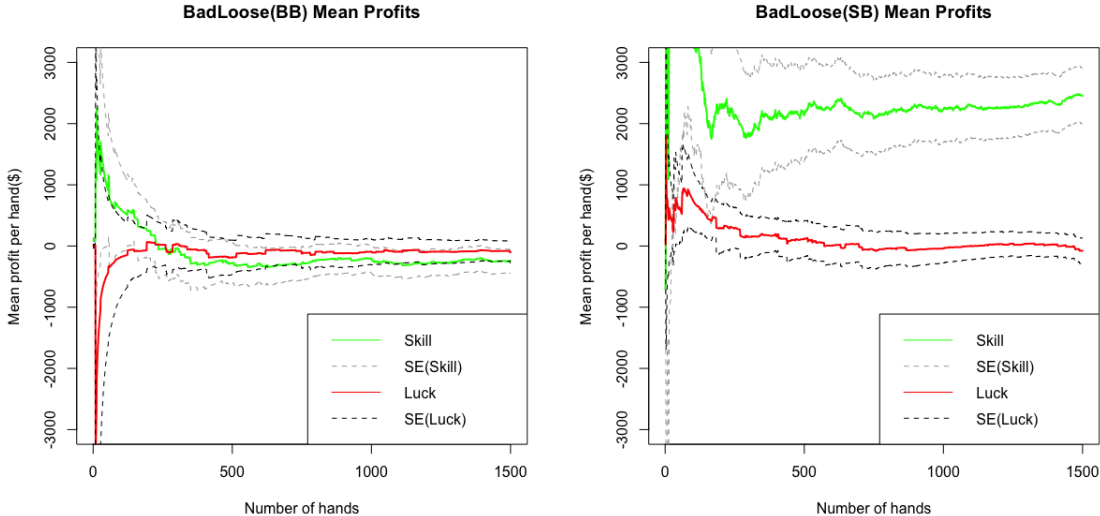


Figure 6: Convergence Plots of BadLoose vs GoodTight (1500 hands each match up).

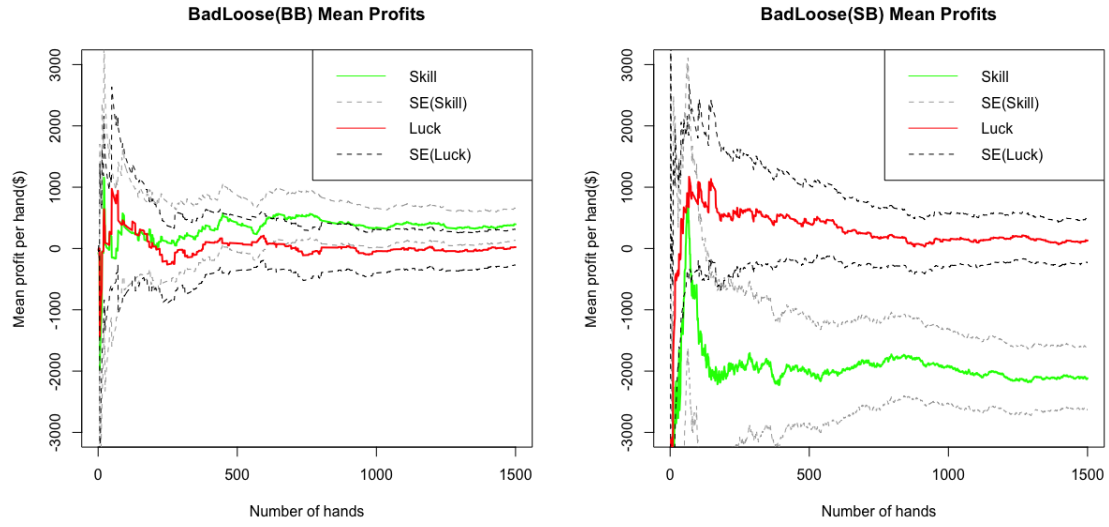


Figure 7: Convergence Plots of BadLoose vs GoodLoose (1500 hands each match up).

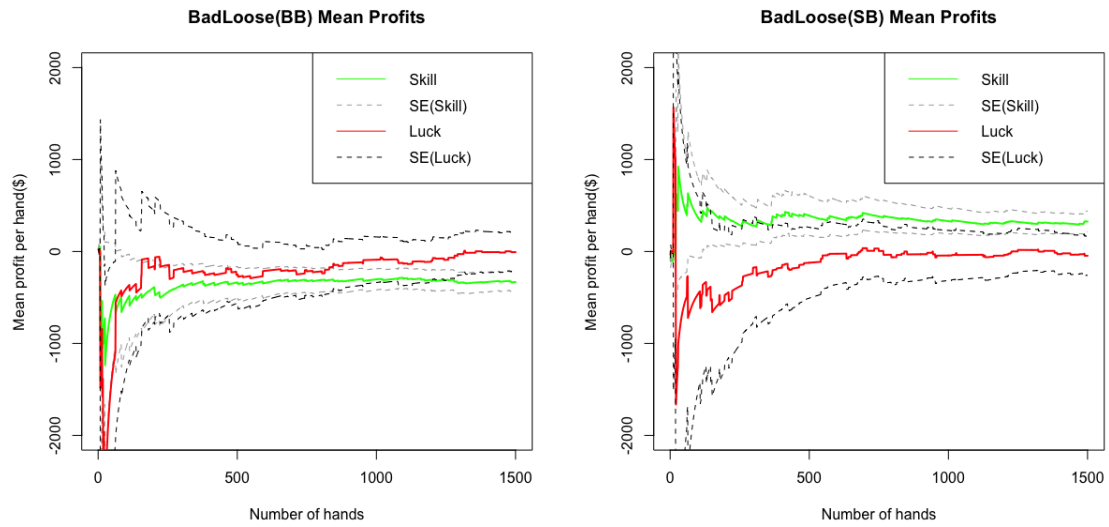


Figure 8: Convergence Plots of BadLoose vs Tommy (1500 hands each match up).

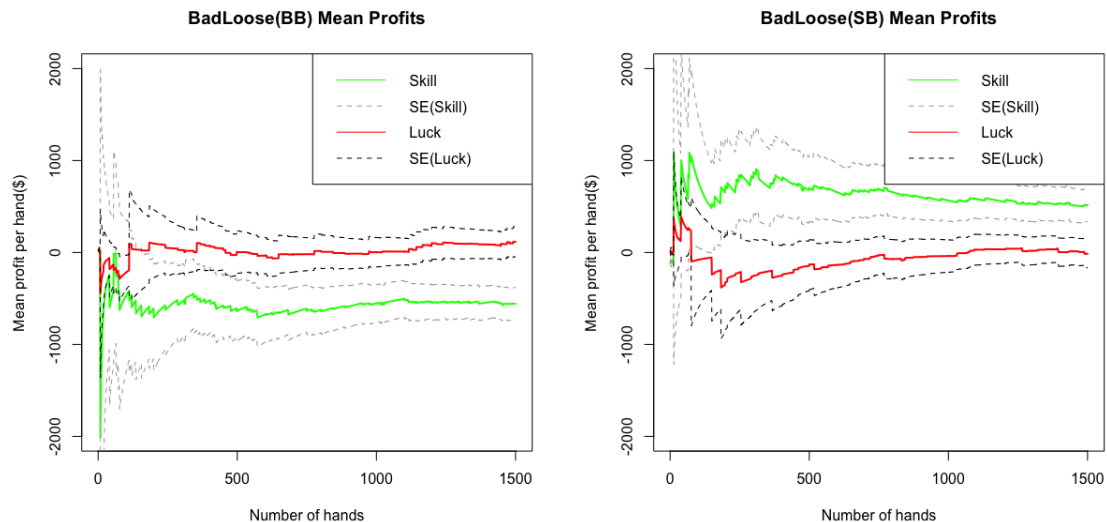


Figure 9: Convergence Plots of BadLoose vs Zelda (1500 hands each match up).

These convergence plots illustrate that as the number of hands increases, the average profit due to luck is indeed converging to something near zero, while the average profit due to skill is usually converging to some other non-zero value which indicates the skill difference between the two players. This provides further evidence in favour of the luck and skill definitions used herein.

7 Conclusion

In this investigation, we completed 21 simulations of 3000 hands each (i.e. 1500 hands in each of the two different table positions), in head-to-head pairings of 6 different player algorithms. The results of the simulations are relatively consistent. The mean expected profit due to skill was larger than the mean expected profit due to luck in 26 out of the 28 cases, and the former was more than 5 times the latter in 19 out of the 28 cases.

Additionally, the correlation between luck and skill had absolute value less than 0.25 in 20 of the 28 cases, indicating that in general there was relatively little correlation between luck and skill.

When developing these definitions of luck and skill, we identified three key criteria by Borm and Van der Genugten [9] for discerning between the two, as well as three possible concerns with the definition from [17]. We will now address these topics in light of the simulation results.

The three discerning criteria:

The skill of a player should be measured as his average result in the long run. While by definition expected profit satisfies this condition, we are also able to note that these results are generally consistent with respect to table positions; in 13 out of the 15 player

combinations, the same player would have a positive skill profit as both big blind and small blind.

It is necessary that expected results vary among players. According to the simulation results, the BadTight player was able to beat the three other new players as well as Tommy and Zelda (the latter only as big blind), while the BadLoose player lost every matchup against the five other algorithms. Furthermore, average expected profits due to skill ranged from as low as 2.71 (GoodTight versus Zelda) to over 2000 (GoodLoose versus BadLoose and GoodTight versus BadLoose), indicating a wide variety results between the players.

The chance elements involved should not prohibit differences to be substantial. Indeed, as previously mentioned, the expected profits due to luck are almost always beaten or dominated by expected profits due to skill. Given that the skill to luck ratios reach as high as 50 to nearly 80, it is evident that luck does not substantially affect the magnitude of results due to skill.

The three possible concerns with the definitions:

Situations can occur where a terrible player may gain expected profit during betting rounds against a better player. While this is still a reasonable concern, there is insufficient evidence to suggest that these situations occur often enough to skew the results. The surprising success of the BadTight player algorithm against the GoodTight algorithm may support this claim, however that may be an artifact how their algorithms interact rather than a systematic issue with the definition.

Luck and skill can be correlated in practice. As noted above, the vast majority of the correlations are weak or non-existent in size. Indeed, the correlation is even negative in 9 of the 28 matchups, and among matches with correlation $|\rho| > 0.25$ the proportion of negative correlations rises to 3 out of 8 (and also the amount of luck is usually very small in those cases). Furthermore, comparing the ratios of the average skill and the average luck, we find that 15 of the 28 combinations had luck and skill profits that were in fact opposite in sign. This appears to be enough evidence to demonstrate that skill and luck are not automatically correlated.

In simple equity calculations the assumption of no future betting or folding can lead to significant overestimates or underestimates of the equity gained/lost from winning/losing the hand. While possible, the simulation results do not indicate any serious problems with the definition due to this shorthand *on average*. Whether because such situations are rare or if they are simply unable to be detected with the current methodology is unclear. Extending the simulations to account for future betting rounds may illuminate this issue.

Summary:

In this investigation, we have evaluated the validity of using expected profits due to dealing and betting actions as measures of luck and skill, respectively. In conclusion, we find that the definitions of luck and skill for Texas Hold'em poker in terms of changes in equity due to card dealing versus bet decisions as in [17] are supported by our simulation study and thus appear to be valid and useful definitions.

We believe that these results have implications beyond mere poker games. In many games, and in many real-life situations, it may be possible to define some notion of “equity” in terms of the conditional expected value of an outcome given the current circumstances. In terms of that definition, intentional actions which affect the equity correspond to skill, while equity changes due to uncontrollable external factors correspond to luck. In that respect, the poker simulation studies considered herein can be seen to provide some early initial guidance about how to distinguish luck from skill in many different aspects of modern society.

References

- [1] DeepStack Web Page. <https://www.deepstack.ai>.
- [2] GitHub Repository (Data Source & R Code). <https://github.com/UofTGuo/Poker>.
- [3] Online Poker Odds Calculator. <http://twodimes.net/poker/>.
- [4] Michael Bowling, Michael Johanson, Neil Burch, and Duane Szafron. Strategy evaluation in extensive games with importance sampling. *Proceedings of the Twenty-Fifth International Conference on Machine Learning*, pages 72–79, 2008.
- [5] Neil Burch, Martin Schmid, Matej Moravčík, and Michael Bowling. AIVAT: A new variance reduction technique for agent evaluation in imperfect information games. *The AAAI-17 Workshop on Computer Poker and Imperfect Information Games*, WS-17-06:304–309, 2016.
- [6] R. Croson, P. Fishman, and D. Pope. Poker superstars: Skill or luck? *Chance*, 21:28, 2008.
- [7] M.A. Dedonno and D.K. Detterman. Poker is a skill. *Gaming Law Review*, 12:31–36, 2008.
- [8] M. Dreef, P. Borm, and B. van der Genugten. Measuring skill in games: several approaches discussed. *Mathematical Methods of Operations Research*, 59:375–391, 2004.
- [9] M. Dreef, P. Borm, and B. van der Genugten. A new relative skill measure for games with chance elements. *Managerial and Decision Economics*, 25(5):255–264, 2004.
- [10] Marc Lanctot, Kevin Waugh, Martin Zinkevich, and Michael Bowling. Monte carlo sampling for regret minimization in extensive games. *Advances in Neural Information Processing Systems*, 22:1078–1086, 2009.
- [11] S. D. Levitt and T. J. Miles. The role of skill versus luck in poker: evidence from the world series of poker. *Journal of Sports Economics*, 15:31–44, 2014.
- [12] S.D. Levitt, T.J. Miles, and A. Rosenfield. Is texas hold ’em a game of chance? a legal and economic analysis. *S 101 Geo. L.J.*, page 581–636, 2012-2013.
- [13] Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisý, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael H. Bowling. Deepstack: Expert-level artificial intelligence in no-limit poker. *CoRR*, abs/1701.01724, 2017.
- [14] J.S. Rosenthal. *Knock on Wood: Luck, Chance, and the Meaning of Everything*. Harper-Collins Canada, 2018.
- [15] Martin Schmid, Neil Burch, Marc Lanctot, Matej Moravcik, Rudolf Kadlec, and Michael Bowling. Variance reduction in Monte Carlo counterfactual regret minimization (vr-mccfr) for extensive form games using baselines. *Preprint*, 2018.

- [16] Frederic Paik Schoenberg. *Introduction to probability with Texas Hold'em examples*. Chapman and Hall, 2011.
- [17] Frederic Paik Schoenberg. Quantification of luck and skill in texas hold'em. *Significance*, 15(6):30–33, 2018.
- [18] Frederic Paik Schoenberg and Yixuan Shao. R Package 'holdem'. <https://cran.r-project.org/web/packages/holdem/holdem.pdf>.
- [19] R.J.D.P. van Loon, M. van den Assem, and D. van Dolder. Beyond chance? the persistence of performance in online poker. *PLoS ONE*, 10:e0115479, 2015.
- [20] Martha White and Michael Bowling. Learning a value analysis tool for agent evaluation. *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 1976–1981, 2009.
- [21] G. Yaari and S. Eisenmann. The hot (invisible?) hand: Can time sequence patterns of success/failure in sports be modeled as repeated random independent trials? *PLOS ONE*, 6:p.e24532, 2011.

Appendix – Definitions of Strategies

GoodTight Player (New Strategy)

Pre-flop: If an AA, AK, KK is drawn, TightPlayer will raise to 4bb, or call anything above 4bb. If QQ, JJ, suited AJ, AQ, KQ, KJ, QJ are drawn, TightPlayer will raise to 3bb, call anything up to 4bb, and fold whenever the amount to call is more than 4bb. For AJ, AQ, KQ, KJ, QJ, 55-TT, TightPlayer will raise to 2bb, call anything up to 3bb, and fold if amount to call is more than 3bb. Any other case, if the amount to call is less than or equal 3bb, TightPlayer will call with probability $bluff = 0.1$ (this value can be set by the user).

Post-flop: When TightPlayer has a pair, he will bet half pot or call up to pot size bets, and fold everything else. For 2 pairs, TightPlayer will bet pot size or call up to 3*pot size bets, and fold everything else. For trips, TightPlayer will bet pot size or call up to 5*pot size bets, and fold everything else. For a made straight or flush, TightPlayer will bet pot size or call up to 5*pot size bets, and fold everything else. For a straight draw (assuming 4 outs only or a 17% or making the straight after the turn and river), TightPlayer will call if amount to call is less than or equal to 17% times the pot + amount to call; for a flush draw (assuming 9 outs) this percentage is 35%. For a full house, four of a kind, or straightflush, TightPlayer will bet pot size and call to everything. For everything else, if the amount to call is less than or equal to the pot size, TightPlayer will call with probability $bluff$.

Turn: When TightPlayer has a pair, he will call up to pot size bets, and fold everything else. For 2 pairs, TightPlayer will bet pot size or call up to 3*pot size bets, and fold everything else. For trips, TightPlayer will bet pot size or call up to 5*pot size bets, and fold everything else. For a made straight or flush, TightPlayer will bet pot size or call up to 5*pot size bets, and fold everything else. For a straight draw, TightPlayer will call if amount to call is less than or equal to 9% times the pot + amount to call; for a flush draw (assuming 9 outs) this percentage is 20%. For a full house, four of a kind, or straightflush, TightPlayer will bet pot size and call to everything. For everything else, if the amount to call is less than or equal to the pot size, TightPlayer will call with probability $bluff$.

River: When TightPlayer has a pair, he will call up to half pot size bets, and fold everything else. For 2 pairs, TightPlayer will bet pot size or call up to 2*pot size bets, and fold everything else. For trips, TightPlayer will bet pot size or call up to 3*pot size bets, and fold everything else. For a made straight, TightPlayer will bet pot size or call up to 4*pot size bets, and fold everything else. For a flush, TightPlayer will bet pot size or call up to 5*pot size bets, and fold everything else. For a full house, four of a kind, or straightflush, TightPlayer will bet pot size and call to everything. For everything else, if there is no betting, TightPlayer will bet half the pot size with $bluff\%$ or raise 1.5 the amount to call with probability $bluff$.

GoodLoose Player (New Strategy)

Pre-flop: If an AA, AK, KK is drawn, TightPlayer will raise to 4bb, or call anything above 4bb. If QQ, JJ, suited AJ, AQ, KQ, KJ, QJ are drawn, TightPlayer will raise to 3bb, call anything up to 4bb, and fold with $(1-bluff=1-0.3)$ whenever the amount to call is more than 4bb. For AJ, AQ, KQ, KJ, QJ, 55-TT, TightPlayer will raise to 2bb, call anything up to 3bb, and fold with $(1-bluff)$ if amount to call is more than 3bb. Any other case, if the amount to call is 1bb, TightPlayer will always call. For any less than 3bb, TightPlayer will call with probability $bluff$.

Post-flop: When TightPlayer has a pair, he will bet half pot or call up to pot size bets, and fold everything else with $(1-bluff)$. For 2 pairs, TightPlayer will bet pot size or call up to 3*pot size bets, and fold everything else with $(1-bluff)$. For trips, TightPlayer will bet pot size or call up to 5*pot size bets, and fold everything else with $(1-bluff)$. For a made straight or flush, TightPlayer will bet pot size or call up to 5*pot size bets, and fold everything else with $(1-bluff)$. For a straight draw (assuming 4 outs), TightPlayer will call if amount to call is less than or equal to 17% times the pot + amount to call and fold with $(1-bluff)$; for a flush draw (assuming 9 outs) this percentage is 35% and fold with $(1-bluff)$. For a full house, four of a kind, or straightflush, TightPlayer will bet pot size and call to everything. For everything else, if the amount to call is less than or equal to 2 times the pot size, TightPlayer will call with probability $bluff$.

Turn: When TightPlayer has a pair, he will call up to pot size bets, and fold everything else with $(1-bluff)$. For 2 pairs, TightPlayer will bet pot size or call up to 3*pot size bets, and fold everything else with $(1-bluff)$. For trips, TightPlayer will bet pot size or call up to 5*pot size bets, and fold everything else with $(1-bluff)$. For a made straight or flush, TightPlayer will bet pot size or call up to 5*pot size bets, and fold everything else with $(1-bluff)$. For a straight draw, TightPlayer will call if amount to call is less than or equal to 9% times the pot + amount to call; for a flush draw (assuming 9 outs) this percentage is 20% and fold with $(1-bluff)$. For a full house, four of a kind, or straightflush, TightPlayer will bet pot size and call to everything. For everything else, if the amount to call is less than or equal to the pot size, TightPlayer will call with probability $0.5*bluff$ and reraise to 2 times the amount to call with probability $0.5*bluff$.

River: Here, whenever TightPlayer decides to fold, he will actually call with probability $0.5*bluff$ and reraise to 2 times the amount to call with probability $0.5*bluff$. When TightPlayer has a pair, he will call up to half pot size bets, and fold everything else. For 2 pairs, TightPlayer will bet pot size or call up to 2*pot size bets, and fold everything else. For trips, TightPlayer will bet pot size or call up to 3*pot size bets, and fold everything else. For a made straight, TightPlayer will bet pot size or call up to 4*pot size bets, and fold everything else. For a flush, TightPlayer will bet pot size or call up to 5*pot size bets, and fold everything else. For a full house, four of a kind, or straightflush, TightPlayer will bet pot size and call to everything. For everything else, if there is no betting, TightPlayer will

bet the pot size with *bluff%*. If there is betting, TightPlayer will call with $0.5 * bluff\%$ or raise 2 times the amount to call with probability $0.5 * bluff$.

Note the new players' strategies does not consider positions.

BadTight Player (New Strategy)

BadTight will go all in preflop only if he gets AA, AK, or KK. Otherwise he will limp in only if amount to call is less than the big blinds. In the three rounds post flop, Badtight will fold or check unless he has made a straight, flush, full house, or quads, in which case he will go all in.

BadLoose Player (New Strategy)

BadLoose will call every bet.

Zelda (Strategy in holdem package)

Pre-flop: AK: Make a big raise if nobody has yet. Otherwise call. AQ: call a small raise, or make one if nobody has yet. AJ, AT, KQ, KJ, QJ: call a tiny raise. A9, KT, K9, QT, JT, T9: call a tiny raise if in late position (within 2 of the dealer). Suited A2-AJ: call a small raise. 22-99: call a small raise. TT-KK: make a huge raise. If someone's raised huge already, then go all in. AA: make a small raise. If there's been a raise already, then double how much it is to you.

Post-flop: If there's a pair on the board and you don't have a set, then check/call up to small bet. Same thing if there's 3-of-a-kind on the board and you don't have a full house or more. If you have top pair or an overpair or two pairs or a set, make a big bet (call any bigger bet). Otherwise, if nobody's made even a small bet yet, then with prob. 20% make a big bluff bet. If you're the last to decide and nobody's bet yet, then increase this prob. to 50%. If you have an inside straight draw or flush draw then make a small bet (call any bigger bet). If you have a straight or better, then just call. Otherwise fold.

Turn: If there's a pair on the board and you don't have a set, then check/call up to small bet. Same thing if there's 3-of-a-kind on the board and you don't have a full house or more. Otherwise, if you have top pair or better, go all in. If you had top pair or over pair but now don't, then check/call a medium bet but fold to more. If you have an inside straight draw or flush draw then check/call a medium bet as well. Otherwise check/fold.

River: If there's a pair on the board and you don't have a set, then check/call up to small bet. Same thing if there's 3-of-a-kind on the board and you don't have a full house or more. Otherwise, if you have two pairs or better, go all in. If you have one pair, then check/call a

small bet. With nothing, go all-in with probability 10%; otherwise check/fold.

Tommy (Strategy in holdem package)

Goes all in with any pocket pair.