Helsinki GSE

HELSINKI GSE DISCUSSION PAPERS 10 · 2023

More skill, less bias?
Breaking down break-even effects in poker

Marco Lambrecht

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HANKEN

Aalto University

Helsinki GSE Discussion Papers

Helsinki GSE Discussion Papers 10 · 2023

Marco Lambrecht:
More skill, less bias? Breaking down break-even effects in poker

ISBN 978-952-7543-09-2 (PDF)
ISSN 2954-1492

Helsinki GSE Discussion Papers:
https://www.helsinkigse.fi/discussion-papers

Helsinki Graduate School of Economics
PO BOX 21210
FI-00076 AALTO
FINLAND

Helsinki, June 2023

More skill, less bias?
Breaking down break-even effects in poker

Marco Lambrecht*

June 6, 2023

Abstract

Chasing and the house money effect are well-known phenomena in dynamic environments of risky activities such as investment decisions or gambling. Recent studies suggest that such behavior emerges from dynamic inconsistency and leads to substantial welfare consequences that can extend beyond financial losses. This study examines novel field evidence from online poker which allows to study biased individuals in a relevant environment where outcomes do not entirely depend on chance. It turns out that individuals exhibiting the house money effect earn less, play less frequently and are of lower relative skill than unbiased individuals. Chasers, on the other hand, earn more, play more frequently and are of higher relative skill than the unbiased group, but chasing is detrimental to their performance and reduces profits by approximately 50%. These findings provide important insights regarding similar environments (such as day-trading) where measuring of individual skill might not be viable.

Keywords: repeated risk-taking, reference-dependence, chasing, skill, poker

JEL-Codes: L83, C72, D81

---

*marco.lambrecht@hanken.fi, Hanken School of Economics & Helsinki GSE
Address: Arkadiankatu 7, PO Box 479, Fi-00101 Helsinki.

I like to thank Mark Dean for valuable discussions within the "Adopt a Paper"-program managed by Daniel Gomez, as well as Aleksandr Alekseev, Benjamin Bushong, Peter Duersch, Sebastian Ebert, Katrin Gödker, Daniel Hauser, Topi Miettinen, Jörg Oechssler, Hannes Rau, Andis Sofianos, Christoph Vanberg, Martin Weber and the seminar audiences at Heidelberg, Helsinki GSE, and FUR Ghent for their helpful comments. Financial support by DFG grant OE 198/5-1 is gratefully acknowledged.

1 Introduction

In many situations, individuals pursue (repeated or continuous) risky activities that require them to decide when to stop. In a relevant subset of such situations, outcomes are not solely dependent on chance. In casino games such as poker or blackjack, sophisticated strategies outperform random play (Bewersdorff, 2021). Similarly, the performance of traders in stock markets depends on identifying profitable investment opportunities. Previous research has shown that individuals continue to take risks after experiencing substantial losses (e.g. Smith et al., 2009; Liu et al., 2010; Suhonen and Saastamoinen, 2018), potentially leading to extreme behavior such as rogue trading which can cost banks substantial amounts of money in fines and provisions - famous examples being Nick Leeson, who lost Baring Banks more than 1 billion USD, Kwenku Adeboli who lost UBS close to 2 billion USD, and Jerome Kerviel who lost Société Général nearly 6 billion USD (Fraser-Mackenzie et al., 2019). A recent study by Heimer et al. (2021) suggests a dynamic inconsistency between planned choices and actual choices in response to experiencing gains and losses and shows that this inconsistency incurs a significant welfare loss. It is a natural question to ask which characteristics of individuals and properties of environments are associated with such behavior. This paper aims to shed light on that question by focusing on the role of skill.

Reference point dependence of preferences is well-known in decision making under risk (see e.g. Kőszegi and Rabin, 2006). Specifically, *chasing behavior* and the *house money effect* have received considerable attention. Exhibiting the house money effect, individuals take more risk when being ahead of their reference point (Thaler and Johnson, 1990). Chasing individuals are willing to increase risk-taking when being behind their reference point, ultimately in pursue to get back to even (Lesieur, 1977). Arguably, online poker is well suited to study such behavior. The environment provides an abundance of data which allows to study individuals in an economically relevant field context with real incentives. Additionally, different implementations of poker feature variation in stake size and to which degree outcomes depend on skill.

In this study, I use a novel poker data-set of mini-tournaments (so called "Sit-and-Go-tournaments"). This data does not suffer from typical confounding drawbacks of other field data.[1] I apply the methodology of Duersch et al. (2020) to measure skill in the different implementations of poker and to approximate individual skill. I analyze how stopping behavior relates to experienced gains and losses and categorize players accordingly.[2] Subsequently, I investigate whether observable

[1]In stock markets, the parameters of recurring investment decisions change over time. For example, there may be rational strategies to increase risk-taking and invest after a market crash that has caused a recent loss. It is difficult to account for such effects (cf. footnote 12 in Zhang and Semmler (2009)). Similar concerns may affect cash game poker data, where the situation at the table changes continuously. For a detailed argument, see section 2.

[2]In the context of my poker data, chasing behavior translates into a smaller likelihood to stop playing when being behind. Analogously, the house money effect is synonymous with a smaller

variables such as profits, frequency of play and relative skill compared to others are associated with specific patterns of stopping. Finally, I study the performance of individuals who have been identified to adopt specific stopping strategies more closely. The results of my analysis on individual level show a remarkable pattern which suggests that success may play an important role in developing behavior that is consistent with chasing. Players exhibiting the house money effect overall earn less, which can be explained by less frequent play and lower individual skill relative to their opponents. On the other hand, chasing individuals earn significantly more than unbiased ones. This is a consequence of more frequent play, as well as higher relative skill on an individual level. At the same time, chasing is detrimental to performance, i.e. individual profits decrease during periods of chasing losses. Regarding the different implementations of poker, I do not find statistically significant differences associated with the amount of skill involved, but chasing is less frequent among high stake players.

Previous papers have studied reference-dependent risk-taking in stock markets (e.g. Barberis et al., 2001; Coval and Shumway, 2005; Zhang and Semmler, 2009; Huang and Chan, 2014) and online poker (Smith et al., 2009; Eil and Lien, 2014) and found evidence for such behavior. Neoclassical economic theory would predict that gains and losses which are relatively small compared to lifetime income should not affect risk aversion. On the other hand, prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) specifies a value function that depends on a reference point and thus provides a framework in which chasing and the house money effect may be observed (Eil and Lien, 2014). Ebert (2020) shows that stopping rules have skewness implications, and skewness seeking individuals should consequently rather opt for behavior consistent with the house money effect instead of chasing.[3] Imas (2016) distinguishes between losses that are realized and losses that only exist "on paper" and shows that, after the latter, individuals become more likely to chase their losses. Likewise, Merkle et al. (2021) derive that individuals are more prone to take on further risks after experiencing gains or losses that are not realized. Factoring in the skill component of poker, it is conceivable that successful players pursue the activity as professionals and, potentially, their stopping behavior may be guided by income targets which they set for themselves when starting to play (see Farber, 2008; Crawford and Meng, 2011; Thakral and Tô, 2021, for structural models of income targeting). Various studies have shown that behavioral biases are more frequent among inexperienced individuals (e.g. List, 2003; Feng and Seasholes, 2005; Locke and Mann, 2005). Fraser-Mackenzie et al. (2019) find that individuals with a profitable history chase less and suggest that higher levels of confidence in own ability might mitigate such behavior. On the other hand, Gervais and Odean (2001) develop a multi-period market model which implies that successful traders are more prone to behavioral biases. Interestingly,

likelihood to stop playing while being ahead.

[3]Note that, using the terminology of financial markets, the house money effect is comparable to a *stop-loss* strategy, while chasing relates to *gain-exit*.

Campbell-Meiklejohn et al. (2008) apply functional magnetic resonance imaging (fMRI) and find that individuals who chase unsuccessfully will chase less in the future. Since better players have higher chances to chase successfully, they are less likely to make such experiences.

The remainder of the paper is structured as follows. The next section describes the data used in this study. Section 3 explains the methodology of the analysis. Section 4 provides the results. Section 5 discusses the findings in a broader context and concludes.

2 Data

Poker was originally played as a five card draw game, but over the years other versions such as Texas Hold'Em have become more popular (Fiedler and Wilcke, 2011). I analyze poker data that was purchased from a commercial vendor (HHSmithy) and was monitored on the online platform "PokerStars" between November 2016 and February 2017. The focus is on *No Limit Texas Hold'Em*. Texas Hold'Em features two private cards and five community cards which are laid out sequentially. The community cards are visible to every player and are common as every player can use them to form a poker hand. Players can look for the strongest combination of five cards out of private and community cards. In No Limit poker, players can freely choose how many chips to bet (conditional on a minimum amount). This includes the option to go "all-in" at every stage of a hand.

In order to analyze the impact of environments on biases, this study analyzes three data sets which are summarized in Table 1. PokerStars offers different speeds of play by varying chip endowments in matches. While *standard* (STD) matches start with 1500 chips per player, *hyper turbo* (HT) matches endow players with 500 starting chips. As the blind structure (i.e., the enforced bets in each hand) of both matches is the same,[4] the smaller amount of starting chips leads to less wiggle room and thus (on average) to fewer hands needed to determine the winner of the match. Note that HT matches also impose a stricter time limit on each decision.[5] Clearly, poker can be played with different stakes. Specifically, data from both micro stake (MS) 3.50$ and high stake (HS) 60$ matches are included in the analysis.

Table 1: Poker data included in this study

	MS	HS
STD	Texas-STD-MS	
HT	Texas-HT-MS	Texas-HT-HS

[4]At the start of the match, the small blind is equal to 10 chips, and the big blind to 20 chips. Subsequently, blinds are increased at fixed intervals.

[5]STD tournaments allow 18 seconds for each decision, while HT tournaments limit this time to 12 seconds. If a player fails to submit an action within that time frame, they will automatically check (if possible) or fold.

The matches in these data sets are so-called *Heads-Up Sit-and-Go tournaments*, which will start whenever two players sit down at the same table. At this point both players pay the entry fee of the match. Then, they are endowed with an equal amount of chips and play until one player has lost all chips to the other player. Subsequently, the winner (i.e., the player holding all the chips) will be rewarded with money worth twice the entry fee.[6]

Previous studies using poker have focused on data from cash game poker. While this has the benefit of utilizing a larger number of observations when interpreting every hand independently, there are some caveats to mention. In cash game poker, the optimal strategy in a given hand depends on the history of previous hands if opponents at the table do not change. On the other hand, a change of opponents (which can happen at any point in time) may have an effect on continuation decisions of the players under consideration. It can even systematically relate to previous outcomes. Some players try to select opponents and might want to leave a table after they won money from a bad player who runs out of funds, while they are even more incentivized to stay in case they lost money to such players. Sit and Go matches, conveniently, do not allow players to join and leave in continuous time. They end when one player won all the chips in play. Leaving a match before it ends would mean that the player gives up, i.e. it is impossible to cash out interim winnings. When starting another match against a new opponent, the optimal strategy to approach the match is arguably independent of any previous match history.

3 Methodology

This section introduces notation and definitions, and describes the methods to derive the main results of this study.

Definition of frequent players

For my analysis, it is important that individuals make conscious decisions to stop at a certain point rather than being forced due to cash constraints. For that reason, I restrict myself to the most frequent players in the data set. I define *most frequent players* as the group of 200 players who have played most frequently within the observed time frame. The choice of this exact number is somewhat arbitrary. There is a trade-off to consider though. By increasing the number, one would gradually add players who have played less matches. This arguably increases the amount of noise. On the other hand, decreasing it would leave less observations for further analysis. I repeat my analysis considering the top 100 most frequent players to verify robustness. The results do not change qualitatively. For the purpose of conciseness, I provide graphs and tables of robustness checks in appendix C.

[6] PokerStars will deduct a small amount of the prize money, the so-called *rake*. This fee amounts to 0.20$ in MS poker and 2.52$ in the HS version.

The reasoning to focus on these individuals is that they have only short intervals between their matches, making it improbable that they have been cash constraint at the end of their previous match.[7] Consequently, I focus my analysis of individual behavior on the most frequent players in each data set. This group may contain rather extreme types of players: on the one hand, "professional" poker players who play in order to earn money, on the other hand potentially "problematic" players (i.e., players who could suffer from gambling addiction). In the light of previous literature (e.g. Bjerg, 2010), it seems worth to point out that these two characteristics are not necessarily mutually exclusive.

Continuation and bracketing

A key ingredient of this study is to identify continuation of play. I define continuation via *sessions*. To fix ideas, consider player i who plays the first match in the data. This starts session s_1^i of player i. If player i plays another match within less than one hour after the first match, I consider both matches to belong to the same session. In this case, player i does not end the session after the first match. On the other hand, if player i does not participate in another match within one hour, session s_1^i ends. The next match then starts a new session, s_2^i, to which the same definition applies. In other words, a session consists of consecutive matches, which are matches with less than one hour between them. A player ends a session whenever the current match is the last match of the session. For robustness, I repeat my analysis for sessions which have less than two hours of breaks in between matches.

The decision to define sessions in this manner is motivated by the study of Eil and Lien (2014), who define continuation based on time between matches. Mostly, the authors concentrate on bracketing of six hours (i.e., assume that players still play the same session when they take breaks of less than six hours). To verify robustness, they compare results to bracketing of four and eight hours. It seems that classifications on individual level stemming from their hazard model are quite sensitive to changes in bracketing. Responding to this, the authors take another approach, which allows that "winnings from an hour or two earlier have already sunk in", finding that "the results of this estimation suggest that many players are bracketing over more recent time frames". In this spirit, I focus my analysis on the definition mentioned above, i.e. one and two hour bracketing. In the following, I will point out when results are sensitive to a switch between the two definitions and refer to appendix C for details. For most of my results, there are no differences. I provide statistics on sessions in the different data sets in table 2. Note that STD matches take about three times as long as HT matches, which might partially explain the differences between the versions.

[7] As my analysis will show, the players under consideration get back to playing in close succession. Yet, it is conceivable that players run out of money on the platform and need to "cash in" additional funds. This procedure is quick and easy (as the platform has incentives to facilitate such transfers) and may be done within seconds. However, the procedure itself might have an effect on the willingness to continue playing (Imas, 2016; Merkle et al., 2021). For further

Table 2: Statistics on matches and sessions - most frequent players

	#Total Sessions	Mean Sessions per Player	Std. Dev. Sessions per player	Mean Matches per Session	Mean Matches per Player	Std. Dev. Matches per player
Texas-STD-MS	10,739	53.7	43.8	3.9	211.4	285.9
Texas-HT-MS	20,768	103.8	68.0	8.1	845.2	808.1
Texas-HT-HS	15,230	76.2	66.4	7.3	557.5	652.6

Note: The table provides information on matches and sessions of the top 200 most frequent players in each data set. Sessions are defined according to bracketing of one hour, i.e. breaks of at least one hour separate sessions from each other.

Identifying biases

Chasing and the house money effect hold that risk preferences are affected by small changes in wealth, i.e. that risk aversion decreases when wealth increases or decreases from the reference point, respectively. Continuous participation in risky activities can be interpreted as additional risk taking and thus increased risk tolerance. Consequently, I focus my analysis on the question whether individuals condition their decision to end a session on current profits. If players fall prey to the house money effect, they will be more likely to continue playing when ahead and rather end a session when they fall behind. Players who chase do exactly the opposite, i.e. they are more likely to continue when they are behind and rather stop when they are ahead. According to that, I classify players into distinct groups.

Specifically, I calculate players' cumulative profits within a session and define a binary variable *behind* to take the value of 0 if profits are positive, and 1 if they are negative.[8] Then, I relate this variable to observations on whether a player ends the session. To fix notation, the variable *end_session* takes the value of 0 whenever a player continues playing, and the value of 1 when a session ends. Then, I perform likelihood-ratio χ^2 tests for each player individually to test for a statistically significant relationship between *end_session* and *behind*.[9] If the observed frequency of *end_session* is small when a player is behind, and if this relationship is significant on a 1% level, I classify them as *chasing*. On the other hand, if the observed frequency of *end_session* is small when a player is ahead, and if this relationship is significant on a 1% level, I classify them as exhibiting the *house money* effect. The remaining individuals are classified as *unbiased*. I complement my analysis with robustness checks for players identified on a 5% level

discussion, see section 5.

[8]When calculating this variable I take the fees charged by the poker platform into account, which is why cumulative profits never sum up to zero in my data.

[9]I opt for likelihood-ratio χ^2 given its robustness with respect to distribution of the data and the relatively small samples under consideration. In simulations, this approach does slightly better at identifying biased individuals than using a hazard model. However, the main results of this paper do not change when identification is based on the latter.

(which I report in appendix C). Since players are heterogeneous with respect to their individual success rates, it is crucial that identification is robust towards this parameter. To ensure this and to understand how sensitive identification is, I run simulations which confirm that the chosen method is appropriate, see appendix B.

Obviously, the counteracting nature of the effects constitutes a problem when both biases are exhibited by the same individual as they might cancel out in the analysis. In order to rule out that the group which I identify as unbiased is stacked with this type of individuals, I examine whether players end their sessions on peaks or troughs (i.e., at maximum or minimum earnings within the current session). It seems fair to assume that those exhibiting both the house money effect and chasing are unlikely to quit when reaching a session high or low. It turns out that the group of players classified as unbiased is significantly more often ending a session at peaks than those classified as exhibiting the house money effect (Mann-Whitney-U test, $z-value = 5.3$). Similarly, unbiased players are more frequently ending sessions at a trough than chasers (Mann-Whitney-U test, $z-value = 8.4$). Thus, the group of unbiased players does not substantially consist of players who exhibit both biases at the same time. For more details, see appendix C.

Since the focus of this study is behavior conditional on being ahead or behind in the current session, it seems worth to mention that players can view their account balance at any point in time. Thus, it is not necessary to remember all outcomes of a potentially long session to understand whether one made profits or lost money - it suffices to remember the amount one had when starting to play. The account balance updates in real time, i.e. entry fees are deducted at the start of a match, and profits are credited immediately after winning a match. Thus, cumulative session profits are salient at all times, preventing players from making mistakes if they condition continuation on this variable.

Variation of environments

To investigate the impact of stake levels on biased behavior, I define the binary variable *high stakes* to take the value of 1 for players of 60$ stakes, and 0 for stakes of 3.50$. Note that the difference is substantial, as there are plenty of stake levels between these two.

Additionally, I measure the heterogeneity of skill involved in the different implementations using the best-fit Elo algorithm (Duersch et al., 2020) - for details, see appendix A. The results show that heterogeneity of playing strengths in poker depends on the version played. Most importantly, outcomes in hyper turbo implementations are significantly more dependent on chance than in standard implementations. For players who are one standard deviation better than their opponent, the win rate in both Texas-HT-MS and Texas-HT-HS is below 51%. If some of the best players compete against some of the worst, they are expected to win close to 55% of their matches. On the other hand, in Texas-STD-MS, one standard deviation of skill difference translates into a winning probability of 53.6%. When the best players compete against the worst, the top players are expected to win more than two thirds of the matches. The disparity between STD and HT versions is confirmed by Levene's test for equality of variances, which shows statistically significant differences. On the other hand, differences between Texas-HT-MS and Texas-HT-HS are not statistically significant. Thus, it seems reasonable to draw the line between standard and hyper turbo poker when distinguishing skill-dependence of environments. For the following analysis, I define the binary variable *higher skill dependence* to take the value of 0 for matches played in hyper turbo settings and 1 in standard settings.

Estimation of individual skill

In order to identify skill on an individual level, it is tempting to look at the (average per game) success rates of players. Yet, this approach could be prone to errors since stopping rules may influence average success rates.[10] Additionally, this perspective ignores the fact that players do not necessarily need to be similar about the type of opponents they match up against. To illustrate this, consider a player who repeatedly competes against the best players and wins about half of the matches. While the success rate does not seem overly impressive, this player should arguably be considered to be of high individual skill. Thus, instead of considering win rates, I use approximations generated from the best-fit Elo algorithm as measures for individual skill.[11] The Elo rating is designed to estimate individual winning probabilities and takes into account how strong the opponent is.

Note that the distributions of Elo ratings are not constant across different versions of poker. Frequent players of the more skill-dependent version overall achieve higher values. Since individual skill is approximated with respect to one specific version, comparisons across different implementations are not recommendable. In particular, the goal is not to claim that certain players of 3.50$ matches are of higher individual skill than high stake players; presumably, rather the opposite is true. Instead, for the analysis, I standardize individual skill by subtracting the average and dividing by the standard deviation of the respective version. I refer to the analysis of this variable as examining *individual skill relative to opponents*, i.e. how players position relative to others within their respective version of poker.

Examining profits "under the influence" of a bias

To analyze whether chasing or the house money effect is in fact a bias in online poker, I estimate the impact of cumulative session profits on results. Specifically,

[10] While stopping rules do not increase or decrease the expected average success rate of an individual, they are consequential for the distribution of average success rates across a group of individuals. For instance, if a group of players follows a *trailing stop loss* strategy (Ebert, 2020), many of them may end up with a low average success rate, while few of them arrive at a very high rate.

[11] In particular, I measure individual ratings using all data from the observation period. In that sense, the measure ignores learning curves during the time frame under consideration. Acknowledging that the data consists of players who play very frequently, it seems fair to assume that their learning curve has flattened out.

I define the binary variable *behind_before_match* to take the value of 1 if cumulative session profits up to the current match are negative, and 0 otherwise. Note that I consider the first match of each session to be included in the group *behind_before_match* = 0. Analogously, I define *ahead_before_match* to take the value of 1 whenever cumulative session profits up to the current match are positive, and 0 otherwise (including the first match of each session). Let won_t^i denote the outcome of the match played by player i at time t, taking the value of 1 if the match was won by player i and 0 otherwise. Then, I estimate effects using linear probability models with standard errors clustered at individual level,

$$won_t^i = \beta_0 + \beta_1 \cdot behind\_before\_match_t^i + \varepsilon_t \tag{1}$$

for the group of chasing individuals, and

$$won_t^i = \beta_0 + \beta_1 \cdot ahead\_before\_match_t^i + \varepsilon_t \tag{2}$$

for the group of individuals exhibiting the house money effect. I complement these estimations with specifications that include control variables for higher skill dependence, high stakes, and current session length. The latter specifically allows to examine the influence of fatigue. To facilitate comparison, I contrast the results of this analysis for the biased groups with those of unbiased players. With this approach I try to understand whether chasing losses or exhibiting the house money effect coincides with changes in individual performance.

Relation between outcome of a session and timing of the next session

Complementing the analysis of continuation behavior in the short-run (i.e. within a session), I take a look at continuation across sessions and how the timing of sessions might relate to outcomes. Let $lost\_session_t^i$ denote the variable which refers to the session of player i which ended at time t, and let it take values of 0 and 1, depending on whether cumulative session profits were positive or not. Furthermore, let $time_t^i$ denote the time that passes between the session ending at time t and the next session of player i. I estimate effects using a linear probability model with standard errors clustered at individual level,

$$time_t^i = \beta_0 + \beta_1 \cdot lost\_session_t^i + \varepsilon_t \tag{3}$$

for each of the three groups. Additionally, I complement these estimations with specifications that include control variables for higher skill dependence, high stakes, and current session length.

4 Results

This section presents the results of this study. The first step is to classify individuals according to the strategy for bias identification laid out in section 3. Table 3 reports the proportion of players who fall into the respective categories. Both biases are observable among the most frequent players of online poker. Notably, chasers outnumber individuals exhibiting the house money effect, except for the high stakes version. The majority of players is classified as unbiased.

Table 3: Individual classification by poker version

	House money	Unbiased	Chasing
Texas-STD-MS	0.110	0.735	0.145
Texas-HT-MS	0.065	0.785	0.150
Texas-HT-HS	0.090	0.845	0.065

Note: The table provides information on player classification on 1% level according to identification laid out in section 3.

As a first step, I systematically analyze bias classification across the different versions. I focus on Texas-STD-MS and Texas-HT-MS to examine the impact of skill-dependent environments. By doing so, the dimension of stakes is held constant. It turns out that there are no differences concerning chasing individuals (likelihood-ratio χ^2 test, $p = 0.991$). However, players exhibiting the house money effect are more frequent in skill-dependent environments. Yet, the result is not statistically significant (likelihood-ratio χ^2 test, $p = 0.102$).[12]

Result 1. *(a) The frequency of chasing individuals does not differ in environments which are more skill-dependent. (b) The increased prevalence of individuals exhibiting the house money effect in more skill-dependent environments is not statistically significant.*

Investigating the link between stake levels and bias classification, I focus on Texas-HT-MS and Texas-HT-HS. This holds the dimension of skill-dependent environments constant, since both versions only differ about stakes. While high stake players do not seem more resistant to the house money effect (likelihood-ratio χ^2 test, $p = 0.506$), there is a stark difference when it comes to the frequency of chasing individuals. In particular, the share of chasers among low stake players is more than twice as large compared to the high stakes group. The result is statistically significant (likelihood-ratio χ^2 test, $p = 0.007$) and robust.[13]

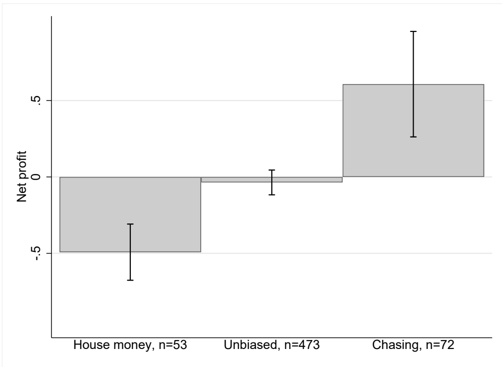
Result 2. *(a) Chasing is less frequent among high stake players. (b) The frequency of individuals exhibiting the house money effect does not differ significantly across stake levels.*

[12] For players identified on 5% level and bracketing of two hours, the different prevalence of players exhibiting the house money effect is statistically significant. For further details on robustness see appendix C.

[13] The difference regarding chasing individuals is significant whenever considering players identified on 1% level. Notably, the share of players exhibiting the house money effect is significantly smaller in Texas-HT-HS when restricting to the top 100 most frequent players. For further details on robustness see appendix C.

The remainder of this section relates the identified groups to other observable variables. I begin with analyzing how classifications are related to overall earnings. Since earnings differ across poker versions, I standardize them for comparison, i.e. I subtract mean earnings and divide by the standard deviation of the respective version. Figure 1 depicts standardized net profits of the different groups. Remarkably, the result for chasing individuals goes in the opposite direction of what one might expect. Despite being viewed as biased, chasers in fact earn significantly more than the unbiased group (Mann-Whitney-U test, $p = 0.000$). Meanwhile, the group of individuals exhibiting the house money effect overall accumulates lower profits (Mann-Whitney-U test, $p = 0.001$).

Figure 1: Relation between *net profits* and bias classification



Note: The graph depicts the average standardized *net profits* in the groups, including 95% confidence intervals.

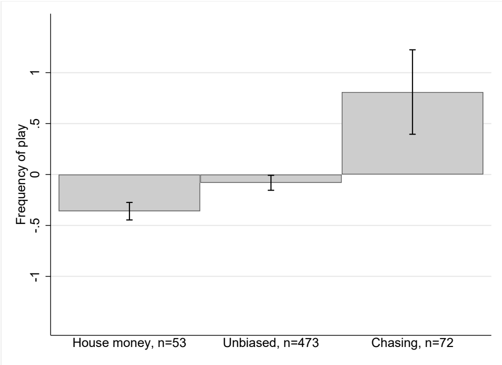
Result 3. *(a) Overall, net profits of chasers are higher. (b) Those exhibiting the house money effect earn less.*

In the following, I report the influence of the variables introduced in section 3 on the house money effect and chasing behavior. With this, I try to unravel the stark differences in net profits across groups. It seems worth to analyze whether the classification of players relates to frequency of play. Players in HT versions tend to play more matches which may be explained by the fact that their matches are shorter. Because of this heterogeneity across different versions, I standardize the number of matches prior to the comparison. The results are depicted in the left panel of figure 2. Notably, there remain stark differences between the groups. Chasing individuals play significantly more matches than unbiased players (Mann-Whitney-U test, $p = 0.000$). However, players exhibiting the house money effect play less (Mann-Whitney-U test, $p = 0.019$). This might seem surprising, given that both biases by definition predict extensive continuation of play.[14]

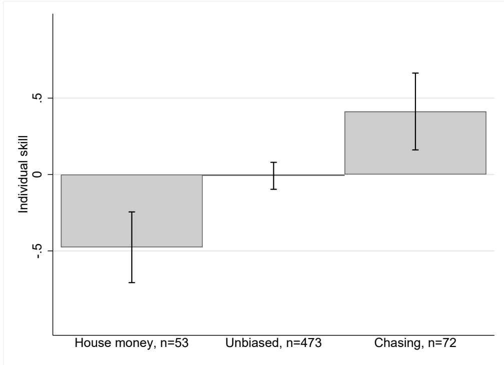
[14]The relation between house money effect and frequency of play becomes statistically insignif-

Figure 2: Relation of biases to *frequency of play* and *individual skill relative to opponents*

(a) Relation between frequency of play and bias classification



(b) Relation between *individual skill relative to opponents* and bias classification



Note: The graphs depict standardized averages within the groups, including 95% confidence intervals.

Result 4. *(a) Chasing individuals play more frequently. (b) Individuals exhibiting the house money effect play less frequently.*

Analyzing relative individual skill, it is worth to note that there are endogenous differences across the data sets. In particular, I measure (on average) higher levels of relative skill in Texas-STD-MS than in HT versions. This might not seem overly surprising given that outcomes in this version are more skill-dependent. To account for heterogeneity, I standardize the variable with respect to the version played before comparing the groups.[15] Figure 2(b) depicts the corresponding results. Overall, players exhibiting the house money effect appear to be of lower individual skill relative to their opponents than unbiased players (Mann-Whitney-U test, $p = 0.001$). Remarkably, there is no evidence that the group of chasers mostly consists of bad players - on the contrary, it is the better ones within their respective version who engage in chasing behavior (Mann-Whitney-U test, $p = 0.002$). Both results are mostly robust with respect to changes of parameters.[16]

Result 5. *(a) The group of chasers consists of players of higher individual skill*

icant under bracketing of two hours when restricting to the top 100 most frequent players (see appendix C). This might be a mechanical issue due to scarcity of observations when focusing on the house money effect among the top 100 most frequent players.

[15] Specifically, I standardize by subtracting the average and dividing by the standard deviation of the respective version.

[16] When considering the top 100 most frequent players and 1 hour bracketing, the differences between chasers and the unbiased group become statistically insignificant. For details, see appendix C.

relative to opponents. (b) Players exhibiting the house money effect are less skilled relative to other players of their respective version.

Table 4: Performance regression - success when behind

	Chasing		Unbiased	
	won	won	won	won
behind before match	-0.01147**	-0.00832*	0.00031	0.00155
	(0.00476)	(0.00434)	(0.00214)	(0.00211)
match in session		-0.00007		-0.00010
		(0.00005)		(0.00007)
high stakes		-0.00727		-0.00239
		(0.00808)		(0.00279)
skill-dependent environment		0.03505***		0.02321***
		(0.00781)		(0.00490)
_cons	0.53752***	0.53001***	0.52152***	0.52081***
	(0.00504)	(0.00508)	(0.00163)	(0.00236)
N	68018	68018	239455	239455
adj. R^2	0.000	0.001	-0.000	0.000

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table provides results of regression specification (1). The first match of each session counts as not being behind.

Tables 4 and 5 illustrate how biases are associated with performance using regression specifications (1) and (2). The focus is on players who are under the influence of their respective bias, and the group of unbiased players serves as control group. It turns out that there are no significant differences in success rates for players exhibiting the house money effect, i.e. they perform neither better nor worse than usual when being ahead. However, the subgroup of house money players in the high stakes version performs significantly worse than their counterparts in the mid stakes versions. Notably, chasing individuals perform worse when they are behind. Taking into account that the poker platform keeps a fee for each match, their expected profits are approximately halved when they are behind before a match. Overall, skill-dependent environments positively alter the proportion of won matches for the subset of players under consideration (i.e., the most frequent players have a larger edge over other players when outcomes are more dependent on skill). This variable is the only significant predictor for the group of unbiased players. The picture does not change when considering individuals identified on 5% level, for details see appendix C.

Result 6. *(a) Chasers perform worse when being behind. (b) Individuals exhibiting the house money effect perform neither better nor worse while being ahead.*

Table 5: Performance regression - success when ahead

	House money		Unbiased	
	won	won	won	won
ahead before match	0.00224	0.00291	-0.00018	0.00053
	(0.00751)	(0.00786)	(0.00221)	(0.00224)
match in session		-0.00005		-0.00010
		(0.00024)		(0.00007)
high stakes		-0.03086**		-0.00236
		(0.01267)		(0.00278)
skill-dependent environment		0.02541*		0.02308***
		(0.01485)		(0.00489)
_cons	0.50116***	0.50265***	0.52174***	0.52128***
	(0.00676)	(0.00700)	(0.00172)	(0.00243)
N	15126	15126	239455	239455
adj. R^2	-0.000	0.001	-0.000	0.000

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table provides results of regression specification (2). The first match of each session counts as not being ahead.

Finally, table 6 relates the outcome of a given session to the time that evolves before a player chooses to start another session. Using regression specification (3), the table provides coefficients on how sessions ending with losses impact the break until the next session for each of the identified groups of players. For the group of unbiased players and the group of players exhibiting the house money effect, a session ending with a loss significantly delays the next session. The coefficient is about twice as large for the latter, with a delay of more than 12 hours. On the other hand, a session ending with a loss does not significantly alter time until the next session for the group of chasing individuals. When adding control variables it turns out that, within each group, high stakes players tend to wait longer until their next session. Similarly, players in Texas-STD-MS take significantly longer breaks except for those in the group of chasing individuals. The length of the previous session does not influence the timing of the next session significantly.

Result 7. *(a) For chasing individuals, the outcome of a session does not influence the timing of the next session. (b) Individuals exhibiting the house money effect take longer breaks after a session that ended with losses. This behavior is similar to the patterns observed among unbiased players.*

Table 6: Regression - time until next session

	House money		Unbiased		Chasing	
	time	time	time	time	time	time
lost session	12.48***	12.38***	6.03***	6.22***	1.19	1.23
	(3.28)	(3.31)	(0.88)	(0.88)	(1.35)	(1.35)
match in session		0.15		0.12		-0.03
		(0.27)		(0.08)		(0.03)
high stakes		27.61***		3.64**		7.81***
		(5.57)		(1.68)		(2.88)
skill-dependent environment		23.74***		17.32***		3.75
		(5.12)		(2.19)		(2.63)
_cons	23.31***	9.39***	20.90***	15.34***	17.29***	15.07***
	(3.03)	(3.43)	(0.80)	(1.13)	(1.27)	(1.44)
N	3214	3214	34437	34437	8020	8020
adj. R^2	0.005	0.026	0.002	0.009	0.000	0.003

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table provides the results of regression specification (3) and session bracketing of one hour. It considers individuals classified on 1% level. Time is measured in hours.

5 Discussion

This paper uses novel field data to investigate reference-dependent risk taking in a context where outcomes are influenced by both skill and chance. I show how chasing and behavior consistent with the house money effect relate to profits, frequency of play and individual level of ability. Moreover, given the skill component of poker, I can directly assess consequences of such behavior on outcomes by comparing individual performance across different points in time. This section summarizes the results and reconciles them with existing literature.

In line with Smith et al. (2009) and Eil and Lien (2014), I find evidence for reference-dependent risk taking among online poker players. Individuals who exhibit behavior consistent with the house money effect overall earn less, are of lower individual skill compared to those who do not exhibit this behavior, and play less frequently. Potentially surprisingly, I find that chasing individuals play more frequently, are of higher individual skill and achieve overall higher profits than the group of unbiased players. At the same time, these players perform worse while they are chasing. This raises the question - what might drive these results?

There is arguably very little indication to suspect that chasing would make individuals successful. But potentially, being successful could make individuals more prone to chase. Gervais and Odean (2001) develop a market model to study overconfidence due to a bias in learning. They consider traders who are uncertain about their own ability and learn over time from observing their successes and failures. In their model, successful traders overestimate their ability and thus become overconfident. When outcomes are not solely dependent on chance, it is conceivable that individuals are uncertain about their ability. Bjerg (2010) relates a higher degree of skill in games to an increased complexity of understanding the issue of control. While fewer people might think that they can win their money back at the roulette table if they just keep playing long enough, this perception could be very different for individuals in an environment where outcomes partly depend on skill. In particular, it may well be the more successful individuals who feel to be in control and are overconfident about their ability to get back to even. This argument is corroborated by the study of Campbell-Meiklejohn et al. (2008) who apply functional magnetic resonance imaging (fMRI) and find that individuals who chase unsuccessfully will chase less in the future. Since better players have higher chances to chase successfully, they are less likely to make such experiences.[17] The conjecture that successful players are more prone to chasing also finds support in the literature about motivated beliefs and reasoning (Bénabou and Tirole, 2016). Better players may attach higher value to holding beliefs about making profits when playing.

Camerer et al. (1997) find that inexperienced taxi drivers set income targets and stop when they reach that target. At the same time, they fail to substitute labor and leisure intertemporally. Farber (2008) develops an empirical model of labor supply and reference-dependent preferences which supports that taxi drivers stop once they reach their reference income level. Yet, this level is subject to unpredictable variation from day to day. The recent study by Thakral and Tô (2021) reconciles previous conflicting interpretations and establishes a structural model in which taxi drivers work toward a reference point that adjusts to deviations from expected earnings. In risky environments where skill plays a role people may pursue the activity as professionals and consider the time spent to be their labor supply. In the given poker data of this paper, chasing individuals play significantly more frequently and achieve higher overall profits. It seems fair to assume that individuals of this group are more likely to consider themselves to be professional poker players. Then, the findings of this paper could potentially indicate that, similar to taxi drivers, professional poker players set themselves income targets and chase until they reach that target. However, the fact that performance differs when being behind speaks against a pure income targeting explanation, as this suggests that strategies within each mini-tournament are adjusted when being behind, i.e., chasing players change their attitude towards risk which in fact is detrimental to their income. The change in performance is relevant, as expected profits are approximately halved while chasing. The result holds when controlling for fatigue. Given the results of Smith et al. (2009) who find that poker players play less

[17] To understand the influence of individual skill on the probability to chase unsuccessfully, I analyze this relationship more closely (see figure 9 in appendix C). It turns out that, assuming a maximum session length of more than 20 matches, an average player is about three times more likely to chase unsuccessfully than a top player. Those who belong to the bottom 5% with respect to individual success rate are about 6 times more likely to chase unsuccessfully than the best players.

cautiously following a loss, it is conceivable that chasing leads to (further) deviation from best-response play which in turn leads to the observed reduction in profits.

The study by Imas (2016) might shed additional light on why players of lower individual skill are less likely to chase. When falling behind during a session, it is possible that a player does not have sufficient funds left available to buy into the next match. While "cashing in" is quick and easy (for example just a couple of mouse clicks if credit card information is stored), the procedure itself might make the difference between paper losses and realized losses and thus explain why risk aversion increases (Imas, 2016). Players who overall earn less or are of lower individual skill relative to others are likely to face this constraint more frequently.

When comparing different level of stakes, I find that chasing behavior is less frequent among high stakes players. In a recent study, Fraser-Mackenzie et al. (2019) analyze characteristics of individuals whose decisions to cease risky activities are affected by strives to break even. They find that traders with a profitable trading history are less affected. Potentially, a profitable trading history could serve as indication for more experience, which is known to reduce biased behavior (List, 2003; Feng and Seasholes, 2005). Arguably, a similar point can be made for the individuals of this study. Palomäki et al. (2013) develop a scale to measure poker player experience. For poker players, it is a common pattern to move up in stakes over time. This makes stakes a suitable proxy for experience. Players who frequently play on micro stakes certainly also have experience, but arguably less compared to frequent high stake players. Moreover, it can not be ruled out that wealth effects play a role. When stakes are high, the outcome of a couple of matches is more relevant. Thus, such effects may contribute to explaining why chasing is less frequent among high stake players.

The findings of this paper add to the recent literature on stopping rules in repeated risk-taking. Ebert (2020) shows how stopping behavior can influence the skewness of a gamble. The house money effect is equivalent to a "stop-loss" strategy and increases skewness, while chasing behavior corresponds to a "gain-exit" strategy and leads to the opposite (Ebert, 2020).[18] Individuals are well-known to be skewness seeking in risky activities (see e.g. Kahneman and Tversky, 1979; Golec and Tamarkin, 1998; Mitton and Vorkink, 2007). This indicates that skilled players do not consciously opt for the distribution of session profits they induce by chasing. Heimer et al. (2021) study dynamic inconsistency which yields deviations from skewness seeking behavior and causes a significant welfare loss. Overall, it is an important observation that particularly successful individuals chase, and it could be an alarming signal for other risky activities (such as day-trading) where precise measuring of individual skill might not be viable.

[18]Indeed, I find that players exhibiting the house money effect tend to have more right-skewed session profits than chasers. Together with the differences in individual skill between the groups, this leads to a negative relation between individual skill and right-skewness of session profits in the data - see figure 10 in appendix C.

References

Barberis, N., M. Huang, and T. Santos (2001): "Prospect theory and asset prices," *The Quarterly Journal of Economics*, 116, 1–53.

Bénabou, R. and J. Tirole (2016): "Mindful economics: The production, consumption, and value of beliefs," *Journal of Economic Perspectives*, 30, 141–164.

Bewersdorff, J. (2021): *Luck, logic, and white lies: The mathematics of games*, CRC Press.

Bjerg, O. (2010): "Problem gambling in poker: Money, rationality and control in a skill-based social game," *International Gambling Studies*, 10, 239–254.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler (1997): "Labor supply of New York City cabdrivers: One day at a time," *The Quarterly Journal of Economics*, 112, 407–441.

Campbell-Meiklejohn, D. K., M. W. Woolrich, R. E. Passingham, and R. D. Rogers (2008): "Knowing when to stop: the brain mechanisms of chasing losses," *Biological Psychiatry*, 63, 293–300.

Coval, J. D. and T. Shumway (2005): "Do behavioral biases affect prices?" *The Journal of Finance*, 60, 1–34.

Crawford, V. P. and J. Meng (2011): "New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income," *American Economic Review*, 101, 1912–32.

Duersch, P., M. Lambrecht, and J. Oechssler (2020): "Measuring skill and chance in games," *European Economic Review*, 127, 103472.

Ebert, S. (2020): "On Taking a Skewed Risk More Than Once," *Available at SSRN 3731565.*

Eil, D. and J. W. Lien (2014): "Staying ahead and getting even: Risk attitudes of experienced poker players," *Games and Economic Behavior*, 87, 50–69.

Farber, H. S. (2008): "Reference-dependent preferences and labor supply: The case of New York City taxi drivers," *American Economic Review*, 98, 1069–82.

Feng, L. and M. S. Seasholes (2005): "Do investor sophistication and trading experience eliminate behavioral biases in financial markets?" *Review of Finance*, 9, 305–351.

Fiedler, I. and A.-C. Wilcke (2011): "The market for online poker," *Available at SSRN 1747646.*

Fraser-Mackenzie, P. A., T. Ma, M.-C. Sung, and J. E. Johnson (2019): "Let's Call it Quits: Break-Even Effects in the Decision to Stop Taking Risks," *Risk Analysis*, 39, 1560–1581.

Gervais, S. and T. Odean (2001): "Learning to be overconfident," *The Review of Financial Studies*, 14, 1–27.

Golec, J. and M. Tamarkin (1998): "Bettors love skewness, not risk, at the horse track," *Journal of Political Economy*, 106, 205–225.

Heimer, R., Z. Iliewa, A. Imas, and M. Weber (2021): "Dynamic inconsistency in risky choice: Evidence from the lab and field," *Available at SSRN 3600583*.

Huang, Y. C. and S. H. Chan (2014): "The house money and break-even effects for different types of traders: Evidence from Taiwan futures markets," *Pacific-Basin Finance Journal*, 26, 1–13.

Imas, A. (2016): "The realization effect: Risk-taking after realized versus paper losses," *American Economic Review*, 106, 2086–2109.

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–291.

Kőszegi, B. and M. Rabin (2006): "A model of reference-dependent preferences," *The Quarterly Journal of Economics*, 121, 1133–1165.

Lesieur, H. R. (1977): *The chase: Career of the compulsive gambler*, Anchor Press.

List, J. A. (2003): "Does market experience eliminate market anomalies?" *The Quarterly Journal of Economics*, 118, 41–71.

Liu, Y.-J., C.-L. Tsai, M.-C. Wang, and N. Zhu (2010): "Prior consequences and subsequent risk taking: New field evidence from the Taiwan Futures Exchange," *Management Science*, 56, 606–620.

Locke, P. R. and S. C. Mann (2005): "Professional trader discipline and trade disposition," *Journal of Financial Economics*, 76, 401–444.

Merkle, C., J. Müller-Dethard, and M. Weber (2021): "Closing a mental account: The realization effect for gains and losses," *Experimental Economics*, 24, 303–329.

Mitton, T. and K. Vorkink (2007): "Equilibrium underdiversification and the preference for skewness," *The Review of Financial Studies*, 20, 1255–1288.

Palomäki, J., M. Laakasuo, and M. Salmela (2013): ""Don't worry, it's just Poker!"-Experience, self-rumination and self-reflection as determinants of decision-making in on-line Poker," *Journal of Gambling Studies*, 29, 491–505.

Siler, K. (2010): "Social and psychological challenges of poker," *Journal of Gambling Studies*, 26, 401–420.

Smith, G., M. Levere, and R. Kurtzman (2009): "Poker player behavior after big wins and big losses," *Management Science*, 55, 1547–1555.

Suhonen, N. and J. Saastamoinen (2018): "How do prior gains and losses affect subsequent risk taking? New evidence from individual-level horse race bets," *Management Science*, 64, 2797–2808.

Thakral, N. and L. T. Tô (2021): "Daily labor supply and adaptive reference points," *American Economic Review*, 111, 2417–2443.

Thaler, R. H. and E. J. Johnson (1990): "Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice," *Management Science*, 36, 643–660.

Tversky, A. and D. Kahneman (1992): "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

Zhang, W. and W. Semmler (2009): "Prospect theory for stock markets: Empirical evidence with time-series data," *Journal of Economic Behavior & Organization*, 72, 835–849.

A Measuring skill and chance

In the following, I describe the method to measure skill and chance in the different data sets of this study. Table 7 summarizes their size. Specifically, it describes them with respect to *regulars*, which are defined as players who have played at least 25 matches.

Table 7: Statistics on matches, players and regulars in the different poker data sets

	#Matches	#Players	#Regulars	Max Matches (Regulars)	Mean Matches (Regulars)	Median Matches (Regulars)
Texas-STD-MS	46,453	14,835	491	2,579	108.0	48
Texas-HT-MS	325,318	38,349	4,575	7,114	110.0	54
Texas-HT-HS	87,322	9,264	782	4,477	175.6	48

In order to measure the degree of chance in the different data sets, I apply the best-fit Elo algorithm as established by Duersch et al. (2020). Due to differences in popularity, the size of the data sets is not balanced. Yet, as Duersch et al. (2020) show in appendix 6.5 of their paper, the algorithm measures skill independent of the size of the data set, conditional on a minimum amount of data to approximate well. The size of the data sets considered in this study arguably fulfill this condition. The procedure to measure skill involves the calibration of the Elo-rating for each data set separately and rating each and every player accordingly.

The Elo-rating approximates playing strengths by assigning a rating to each player. It is constructed to achieve the calculation of expected winning probabilities whenever two players meet in a competition at time t,

$$E_{ij}^t := \frac{1}{1 + 10^{-\frac{R_i^t - R_j^t}{400}}}.$$

The rating R_i^t of player i is an empirical measure of player i's playing strength. More specifically, player i's chance of winning against j is dependent on the difference in ratings via the expected score $E_{ij}^t \in (0,1)$, which can also be thought of as i's expected payoff (e.g. when a draw is counted as $\frac{1}{2}$). The Elo ratings of the players i, j who are in match t are updated as follows,

$$R_i^{t+1} = R_i^t + k \cdot (S_{ij}^t - E_{ij}^t).$$

Here, S_{ij}^t denotes the observed score of player i in match t, which takes the value of 0 if player i lost the match, and 1 in case the match was won.[19] The ratings of players who are not involved in match t do not change. The best-fit Elo algorithm calibrates the parameter k for each data set individually. Changing the parameter affects the updates of ratings, and consequently also future expected scores. I indicate this

[19]Note that the Elo-rating is designed for situations where $S_{ij}^t \in [0,1]$ and $S_{ij}^t + S_{ji}^t = 1$.

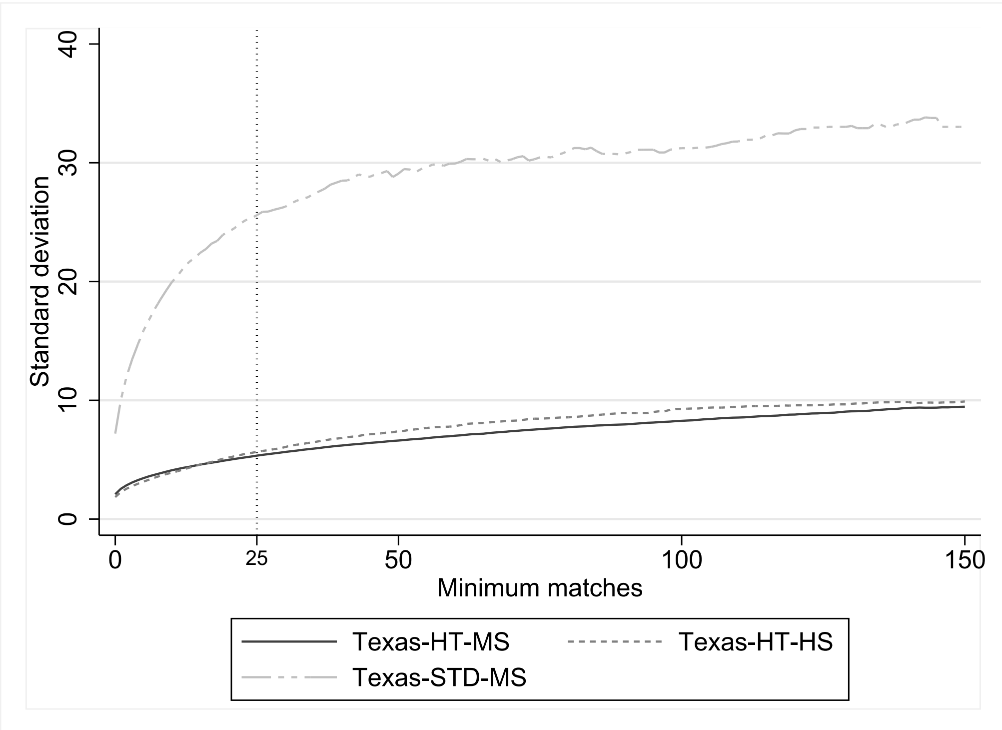
dependency by $E_{ij}^t(k)$. In order to achieve the best possible calibration, the optimal value k^* is chosen as:

$$k^* := \arg\min_k \frac{1}{T} \sum_{t \in T} \left(S_{ij}^t - E_{ij}^t(k) \right)^2$$

Every match t results in two error terms, one for each player competing in match t. Intuitively speaking, k^* is chosen so that prediction errors (ex post) are minimized.

Following Duersch et al. (2020), I focus the analysis on "regulars", which are players who have played at least 25 matches in the data set. Due to the updating nature of Elo ratings, initial ratings might not approximate true playing strengths well. Figure 3 depicts the standard deviation of ratings measured by the best-fit Elo algorithm, conditional on the threshold of minimum matches for players to be included in the data set. The cutoff of 25 matches seems reasonable, since the curves flatten out at this point.

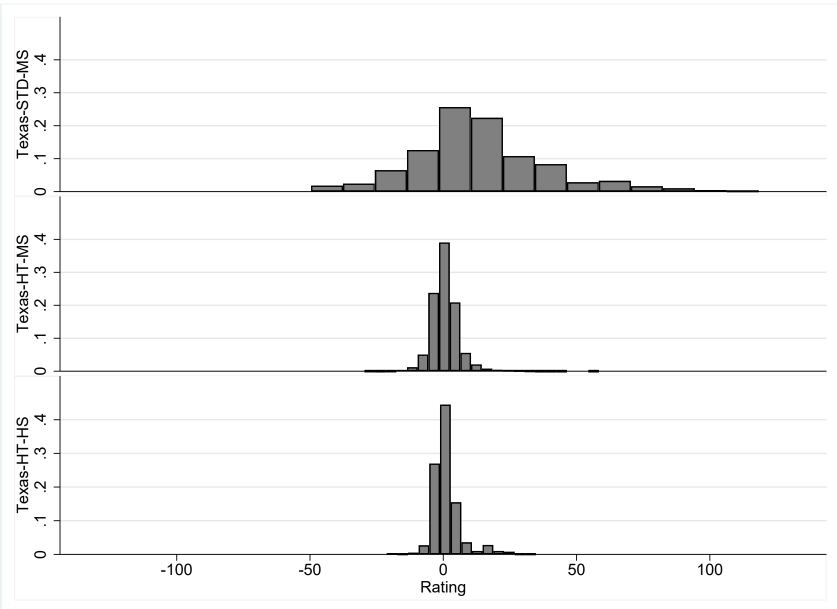
Figure 3: Standard deviation of rating distributions for different cut-off values



Note: Minimum matches refers to the threshold of matches per player in the data set to be included in the calculation of the standard deviation. The vertical dotted line indicates a minimum of 25 matches.

Once all players are rated according to the best-fit Elo algorithm, the focus is on the standard deviation of the rating distributions of each game. In the Elo rating, a given difference in ratings of two players corresponds directly to the winning probabilities when the two players are matched against each other. Thus, the more heterogeneous the ratings are, the better one can predict the winner of a match. If the distribution of Elo ratings is very narrow, then even the best players are not predicted to have a winning probability much higher than 50%. The wider the distribution, the more likely are highly ranked players to win when playing against lowly ranked players, and the more heterogeneous are the player strengths. In my data, the rating distributions of all games are unimodal, see figure 4. Thus, it is possible to interpret the standard deviation of ratings as a measure of skill. It is clearly observable that the STD version of Texas Hold'em has a much wider distribution than the HT versions. For further details on the best-fit Elo algorithm, see Duersch et al. (2020).

Figure 4: Rating distributions for different versions of poker



Note: The rating distributions are centered close to zero by design. Only regulars, i.e. players who have completed at least 25 matches within the data set, are included.

Table 8 reports the main result of the algorithm. Note that the focus of measurement is the standard deviation of Elo rating distributions of regular players. The table reports the minimum and maximum rating, and the rating of the 1% and the 99% percentile player. One can transform the standard deviation of each game into the corresponding winning probability of a player who is exactly one standard deviation better than his opponent. This probability is denoted as p^{sd}. For comparison, the table also provides the winning probablities when a 99% percentile player is matched against a 1% percentile player, which is denoted as p_1^{99}. The winning probability p_1^{99} can be used to calculate the number of matches necessary so that a player who is in the top percentile wins more than half of the matches with a probability larger than 75% against an opponent that is in the bottom percentile. This number is reported in the Rep_1^{99} column. Analogously, the Rep_{sd} column reports the respective number of matches necessary for a player who is one standard deviation better than the opponent.

Table 8: Summary statistics on the distribution of Elo ratings

	Std Dev	Min	1%	99%	Max	p_{sd}	p_1^{99}	Rep_1^{99}	Rep_{sd}
Texas-STD-MS	25.3	-49.6	-44.5	91.9	107.4	53.6	68.7	3	89
Texas-HT-MS	5.3	-29.6	-11.0	18.9	54.6	50.8	54.3	60	1,777
Texas-HT-HS	5.6	-21.2	-9.1	23.2	32.5	50.8	54.6	55	1,777

Note: In contrast to chess, ratings are centered on zero by design. Only regulars, i.e. players who played at least 25 matches, are included in this statistic.

Comparing the different versions of poker, it turns out that STD matches involve significantly more skill than HT matches. Switching from Texas-STD-MS to Texas-HT-MS matches decreases the winning probability p_1^{99} from 66.9% to 54.4%. Regarding different stake levels, Texas-HT-MS has a slightly wider distribution of ratings than Texas-HT-HS. The differences in the standard devations of STD and HT distributions are highly statistically significant, see table 9.

Table 9: Statistical test for equality of variances of rating distributions

	M0	M50	M10
Texas-STD-MS vs. Texas-HT-MS	0.000	0.000	0.000
Texas-STD-MS vs. Texas-HT-HS	0.000	0.000	0.000
Texas-HT-MS vs. Texas-HT-HS	0.173	0.992	0.902

Note: p-values of Levene's Test centered at the mean (M0), at the median (M50), and using the 10% trimmed mean (M10).

It might be surprising that Texas-HT-HS shows a similar heterogeneity of playing strengths as Texas-HT-MS, given that higher stakes should attract better (and thus, a larger variety of) players. However, Siler (2010) finds that the heterogeneity of play styles decreases when comparing high stakes to low stakes. Specifically, high stakes players seem to converge towards more successful play styles, thus increasing the influence of chance on outcomes due to the similarity of their strategies. This seems to counteract increases in heterogeneity from potentially attracting a wider range players.

It is worth to point out that the goal of this paper is not to measure the overall amount of skill in poker (for which it might be questionable to separate STD and HT implementations, or different levels of stakes). However, the focus of this study is to include the dimension of *skill-dependent environments* for systematic comparisons. For that matter, I confirm that variation in speed of play influences the measured degree of skill in the expected direction. When reducing starting chips and putting time pressure on decisions, many matches are decided after only a few hands, often including all-in situations before any community card is revealed. This, by design, increases the influence of chance on outcomes.

B Bias identification simulations

This section provides the result of simulations regarding the identification method described in section 3 of the paper. The objective of this exercise is two-fold. On the one hand, it serves the purpose to ensure that the method works accurately. Note that the poker players in the observed data differ with respect to their individual win rate. One might be concerned that this heterogeneity in success rates might systematically relate to identification and thus mechanically explain some of the results. It turns out that this is not the case. Additionally, the simulations facilitate to understand how sensitive identification is.

The goal of the identification strategy of section 3 is to recognize individuals that exhibit behavior consistent with *chasing* or the *house money* effect. Both effects describe changes in risk-taking behavior dependent on ones position in wealth with respect to the reference point. In the context of risky environments, continuation of the risky activity can be interpreted as additional risk-taking. The natural reference point for wealth is the wealth level when starting the risky activity. Thus, chasing behavior translates into a larger probability to continue the risky activity after losing money. The house money effect, on the contrary, translates into a larger probability to continue when an individual gained money.

To model these patterns, I simulate data of (artificial) poker players who play according to a stochastic process. After each match, the player continues to play with a probability that depends on whether profits are positive or negative compared to the start of the session. In particular, the probability to continue when being ahead, p_a, may differ from the probability to continue when being behind, p_b, which then relates to the behavioral patterns described above.

$$(p_a, p_b) \in \{(0.85, 0.65), (0.8, 0.7), (0.75, 0.75), (0.7, 0.8), (0.65, 0.85)\}.$$

A player for whom $p_a > p_b$ behaves consistent with the house money effect. Those who are more likely to continue when being behind, $p_a < p_b$, are chasing. When $p_a = p_b$, the continuation behavior of the player is unbiased. Furthermore, I consider players who vary with respect to their individual success rates, i.e. their probability to win a match, $p_w \in \{0.45, 0.525, 0.6\}$. These values correspond to the winning probabilities of the 5%, 50% and 95% percentile players in the observed data. The cumulative session profit which determines whether a player is ahead or behind is calculated analogously to the data of the main paper (including the fee taken by the poker platform). Continuation of a session is determined randomly according to cumulative profits and the respective continuation probability, up to a maximum length of 40 matches in one session. Overall, play is repeated for a total of 100 sessions for each player. The simulations consist of 10,000 players each.

Table 10: Simulations for bias identification on 1% level

Winning %	Continuation % ahead	Continuation % behind	House money effect %	Unclassified %	Chasing %
45.0	85.0	65.0	96.55	3.45	0.00
45.0	80.0	70.0	35.48	64.52	0.00
45.0	75.0	75.0	0.58	98.85	0.57
45.0	70.0	80.0	0.00	64.89	35.11
45.0	65.0	85.0	0.00	1.95	98.05
52.5	85.0	65.0	98.32	1.68	0.00
52.5	80.0	70.0	38.77	61.23	0.01
52.5	75.0	75.0	0.49	98.94	0.57
52.5	70.0	80.0	0.00	59.41	40.59
52.5	65.0	85.0	0.00	1.09	98.91
60.0	85.0	65.0	98.76	1.24	0.00
60.0	80.0	70.0	39.41	60.59	0.00
60.0	75.0	75.0	0.47	99.00	0.53
60.0	70.0	80.0	0.00	61.30	38.70
60.0	65.0	85.0	0.00	1.77	98.23

Note: The table provides the results of simulations of 10,000 players and 100 sessions per player. The maximum length of a session is 40 matches.

The generated data is then used to test the identification method applied to the observed data of the main paper. Again, the binary variable *behind* describes whether cumulative profits within a session are negative. The variable *end_session* takes the value of 0 whenever a player continues playing, and the value of 1 when a session ends. I perform likelihood-ratio χ^2 tests for each player to test for a statistically significant relationship between *end_session* and *behind*. If the observed frequency of *end_session* is smaller than the expected frequency when a player is behind, and if this relationship is significant on a 1% level, I classify them as *chasing*. On the other hand, if the observed frequency of *end_session* is smaller than the expected frequency when a player is ahead, and if this relationship is significant on 1% level, I classify them as exhibiting the *house money* effect. The remaining individuals are classified as *unbiased*. I complement this with player classifications on 5% level. Tables 10 and 11 depict the results.

It turns out that a difference of 10 percent in the probabilities p_a and p_b suffices to correctly identify more than one third of players on 1% level, and close to two

Table 11: Simulations for bias identification on 5% level

Winning %	Continuation % ahead	Continuation % behind	House money effect %	Unclassified %	Chasing %
45.0	85.0	65.0	99.14	0.86	0.00
45.0	80.0	70.0	58.74	41.26	0.00
45.0	75.0	75.0	2.52	94.88	2.60
45.0	70.0	80.0	0.00	40.26	59.74
45.0	65.0	85.0	0.00	0.29	99.71
52.5	85.0	65.0	99.65	0.35	0.00
52.5	80.0	70.0	62.88	37.11	0.01
52.5	75.0	75.0	2.33	94.96	2.71
52.5	70.0	80.0	0.00	34.99	65.01
52.5	65.0	85.0	0.00	0.16	99.84
60.0	85.0	65.0	99.80	0.20	0.00
60.0	80.0	70.0	63.34	36.66	0.00
60.0	75.0	75.0	2.32	95.23	2.45
60.0	70.0	80.0	0.00	37.03	62.97
60.0	65.0	85.0	0.00	0.34	99.66

Note: The table provides the results of simulations of 10,000 players and 100 sessions per player. The maximum length of a session is 40 matches.

thirds on 5% level. A false classification in the opposite direction is highly unlikely. When the difference in continuation probabilities is 20 percent, nearly all players are classified correctly. As expected, identification errors are approximately 1% for $(p_a, p_b) = (0.75, 0.75)$ when players are classified on 1% level, and approximately 5% when players are classified on 5% level.

C Robustness

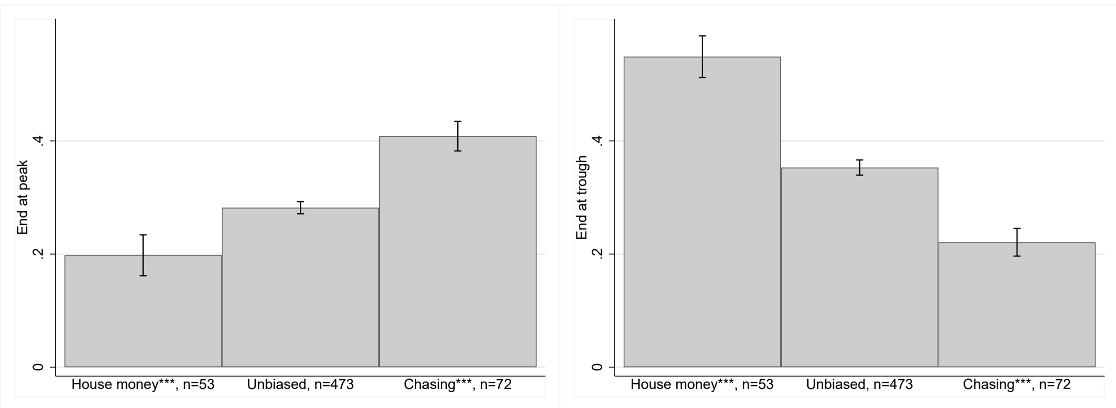
This section provides robustness checks of the analysis of the main paper. Figure 5 depicts the average rates for individuals of different groups to end sessions at a high or a low point of their current session. It serves the purpose to verify that the group of unbiased players differs significantly from both the group of individuals exhibiting the house money effect and the group of chasers. In particular, unbiased players are much more likely to end a session at a peak (i.e. the highest profit of the current session) compared to those exhibiting the house money effect. At the same time, they are more likely to end sessions at a trough (i.e. the lowest profit of the current session) compared to chasers. This indicates that the group of unbiased players does not substantially consist of actors who exhibit both biases at the same time.

Furthermore, I provide robustness checks for individual classification according to the identification strategy laid out in section 3. In particular, I provide results for sessions being defined to allow for breaks of up to two hours between subsequent matches. I also include results for the top 100 most frequent players, as well as identification on 5% level. Table 12 summarizes statistics on sessions and matches for the different parameters. Tables 13, 14, 15, 16, 17, 18 and figures 6, 7, 8 provide the robustness checks for the respective results of the main paper. Tables 19 and 20 show the regressions on performance for identification on 5% level, and table 21 provides the regressions on time until the next session for identification on 5% level. Overall, results are essentially identical to those based on the specification of the main paper.

Figures 9 and 10 depict results that corroborate the discussion of the main paper by providing evidence on the likelihood to chase unsuccessfully and on the relation between individual skill and right-skewness of session profits.

Figure 5: Ending sessions at highs or lows

(a) Ending at peaks (b) Ending at troughs



Note: The graph depicts average individual shares of ending a session at a peak or a trough by group classification, including 95% confidence intervals.

Table 12: Statistics on matches and sessions - most frequent players

	#Total Sessions	Mean Sessions per Player	Std. Dev. Sessions per player	Mean Matches per Session	Mean Matches per Player	Std. Dev. Matches per player
Texas-STD-MS	7,645	76.5	50.7	4.5	342.8	359.4
Texas-HT-MS	12,758	127.6	76.4	9.6	1,222.5	1,010.5
Texas-HT-HS	11,780	117.8	70.2	8.2	967.5	717.2
Texas-STD-MS	5,746	57.5	34.3	7.6	342.8	359.4
Texas-HT-MS	9,312	93.1	50.5	16.1	1,222.5	1,010.5
Texas-HT-HS	8,309	83.1	46.0	13.2	967.5	717.2
Texas-STD-MS	8,417	42.1	30.4	6.4	211.4	285.9
Texas-HT-MS	15,556	77.8	47.1	13.4	845.2	808.1
Texas-HT-HS	11,133	55.7	44.3	11.7	557.5	652.6

Note: The table provides information on sessions and matches, depending on different parameters. The top section refers to the top 100 most frequent players, and sessions defined according to bracketing of one hour. The mid section refers to the top 100 most frequent players, and sessions defined according to bracketing of two hours. Finally, the bottom section refers to the top 200 most frequent players, and sessions defined according to bracketing of two hours.

Table 13: Individual classification by poker version

	House money	Unbiased	Chasing
Texas-STD-MS	35	135	38
Texas-HT-MS	25	130	45
Texas-HT-HS	31	142	27
Total	91	397	110
Texas-STD-MS	20	157	23
Texas-HT-MS	12	166	22
Texas-HT-HS	16	174	10
Total	48	497	55
Texas-STD-MS	36	130	34
Texas-HT-MS	20	152	28
Texas-HT-HS	26	154	20
Total	82	436	82

Note: The table provides information on player classification of the top 200 most frequent players according to the identification strategy laid out in section 3. The top section refers to identification on 5% level, and sessions defined according to bracketing of one hour. The mid section refers to identification on 1% level, and sessions defined according to bracketing of two hours. Finally, the bottom section refers to identification on 5% level, and sessions defined according to bracketing of two hours.

Table 14: Bias classification and *skill-dependent environments*

	1% classification				5% classification			
	1 hour bracketing		2 hour bracketing		1 hour bracketing		2 hour bracketing	
	top 100	top 200	top 100	top 200	top 100	top 200	top 100	top 200
chasing losses	0.639	0.911	0.615	0.753	0.710	0.608	0.304	0.212
house money effect	0.406	0.102	0.804	0.132	0.762	0.193	0.285	**0.012**

Note: p-values of likelihood-ratio χ^2 tests. The top row shows the comparison of the shares of players chasing losses to the share of unbiased players across Texas-STD-MS and Texas-HT-MS (i.e., poker versions that differ along the dimension of skill-dependence, while stakes are held constant). The bottom row shows the comparison of the shares of players exhibiting the house money effect to the share of unbiased players, respectively. While most comparisons do not show statistically significant differences, the share of house money effect players is larger in Texas-STD-MS than in Texas-HT-MS for one specification.

Table 15: Bias classification and *high stakes*

	1% classification				5% classification			
	1 hour bracketing		2 hour bracketing		1 hour bracketing		2 hour bracketing	
	top 100	top 200	top 100	top 200	top 100	top 200	top 100	top 200
chasing losses	**0.025**	**0.007**	**0.069**	**0.029**	0.273	**0.026**	0.205	0.263
house money effect	**0.050**	0.506	**0.003**	0.543	**0.080**	0.667	**0.000**	0.433

Note: p-values of likelihood-ratio χ^2 tests. The top row shows the comparison of the shares of players chasing losses to the share of unbiased players across Texas-HT-MS and Texas-HT-HS (i.e., poker versions that differ along the dimension of stakes, while skill-dependence is held constant). The bottom row shows the comparison of the shares of players exhibiting the house money effect to the share of unbiased players, respectively. The share of players identified to chase losses on 1%-level is smaller in Texas-HT-HS for all specifications, as well as one specification using 5%-level identification. On the other hand, the share of players exhibiting the house money effect is smaller in Texas-HT-HS whenever restricting to the top 100 most frequent players.

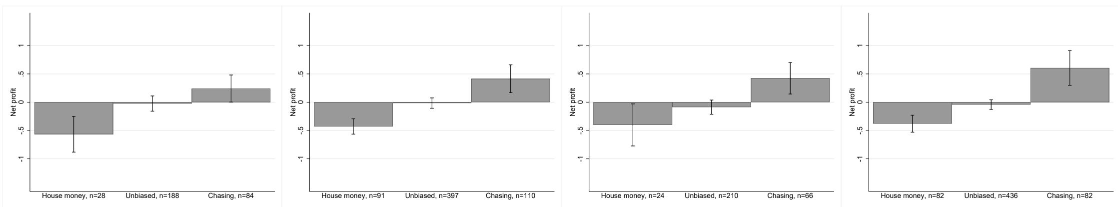
Table 16: Bias classification and *net profits*

	1% classification				5% classification			
	1 hour bracketing		2 hour bracketing		1 hour bracketing		2 hour bracketing	
	top 100	top 200	top 100	top 200	top 100	top 200	top 100	top 200
chasing losses	**0.058**	**0.000**	**0.002**	**0.000**	0.211	**0.003**	**0.005**	**0.000**
house money effect	**0.003**	**0.001**	**0.015**	**0.000**	**0.006**	**0.000**	**0.092**	**0.000**

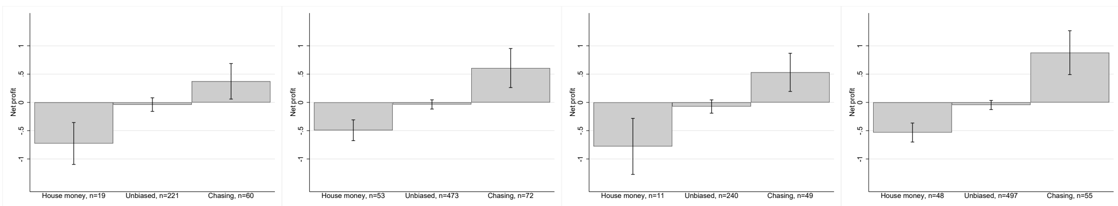
Note: p-values of exact Mann-Whitney-U tests. Net profits are standardized with respect to the particular poker implementation. The top row shows the comparison of profits of players chasing losses to the profits of unbiased players. The bottom row shows the comparison of profits of players exhibiting the house money effect to the profits of unbiased players, respectively.

Figure 6: Bias classification and *net profits*

(a) 1 hour, top 100 (b) 1 hour, top 200 (c) 2 hour, top 100 (d) 2 hour, top 200



(e) 1 hour, top 100 (f) 1 hour, top 200 (g) 2 hour, top 100 (h) 2 hour, top 200



Note: The graph depicts the average standardized *net profits* in the groups, including 95% confidence intervals. The top row depicts groups identified on 5% level, the bottom row those identified on 1% level. The parameters of the analysis vary across the different figures (i.e., bracketing of 1 hour versus 2 hours as well as the top 100 most frequent players versus the top 200 most frequent players of each version).

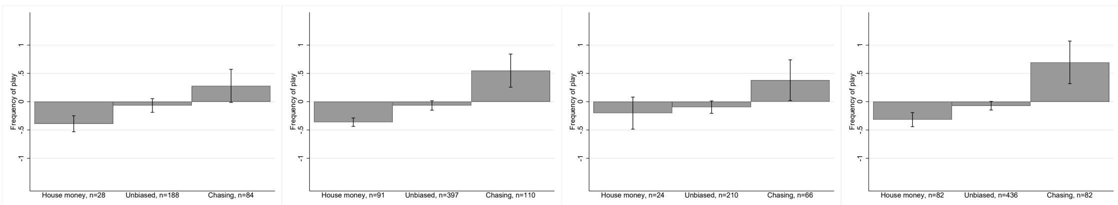
Table 17: Bias classification and *frequency of play*

	1% classification				5% classification			
	1 hour bracketing		2 hour bracketing		1 hour bracketing		2 hour bracketing	
	top 100	top 200	top 100	top 200	top 100	top 200	top 100	top 200
chasing losses	**0.004**	**0.000**	**0.001**	**0.000**	**0.064**	**0.000**	**0.017**	**0.000**
house money effect	**0.059**	**0.019**	0.696	**0.001**	**0.030**	**0.001**	0.553	**0.000**

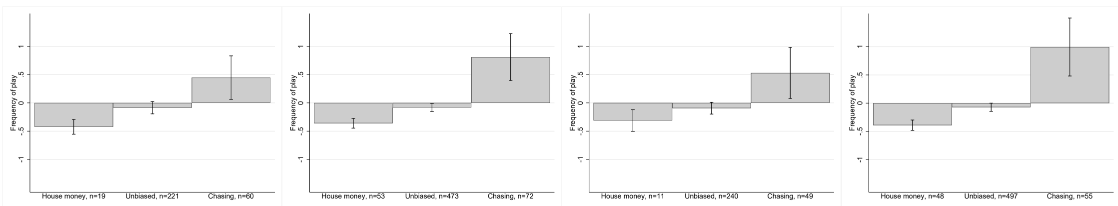
Note: p-values of exact Mann-Whitney-U tests. Frequency of play is standardized with respect to the particular poker implementation. The top row shows the comparison of playing frequency of players chasing losses to the profits of unbiased players. The bottom row shows the comparison of playing frequency of players exhibiting the house money effect to the profits of unbiased players, respectively.

Figure 7: Bias classification and *frequency of play*

(a) 1 hour, top 100 (b) 1 hour, top 200 (c) 2 hour, top 100 (d) 2 hour, top 200



(e) 1 hour, top 100 (f) 1 hour, top 200 (g) 2 hour, top 100 (h) 2 hour, top 200



Note: The graph depicts the average standardized frequency of play in the groups, including 95% confidence intervals. The top row depicts groups identified on 5% level, the bottom row those identified on 1% level. The parameters of the analysis vary across the different figures (i.e., bracketing of 1 hour versus 2 hours as well as the top 100 most frequent players versus the top 200 most frequent players of each version).

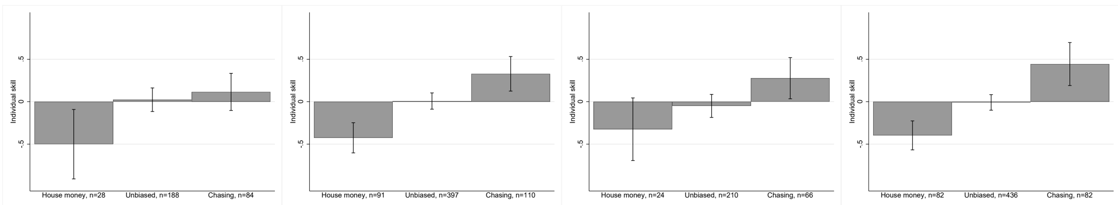
Table 18: Bias classification and *individual skill relative to opponents*

	1% classification				5% classification			
	1 hour bracketing		2 hour bracketing		1 hour bracketing		2 hour bracketing	
	top 100	top 200	top 100	top 200	top 100	top 200	top 100	top 200
chasing losses	0.467	**0.002**	**0.037**	**0.000**	0.605	**0.005**	**0.031**	**0.001**
house money effect	**0.007**	**0.001**	**0.015**	**0.000**	**0.014**	**0.000**	0.252	**0.001**

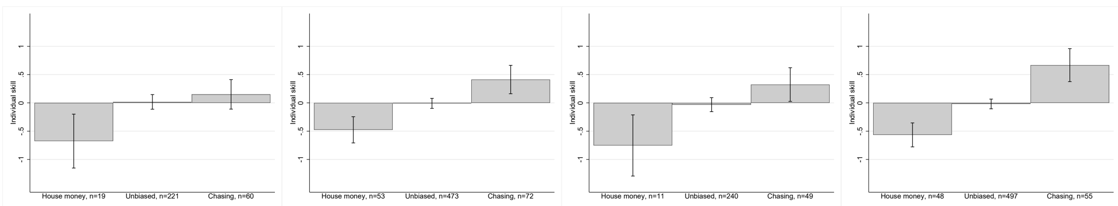
Note: p-values of exact Mann-Whitney-U tests. Individual skill relative to opponents is standardized with respect to the particular poker implementation. The top row shows the comparison of individual skill of players chasing losses to the profits of unbiased players. The bottom row shows the comparison of individual skill of players exhibiting the house money effect to the profits of unbiased players, respectively.

Figure 8: Bias classification and *individual skill relative to opponents*

(a) 1 hour, top 100 (b) 1 hour, top 200 (c) 2 hour, top 100 (d) 2 hour, top 200



(e) 1 hour, top 100 (f) 1 hour, top 200 (g) 2 hour, top 100 (h) 2 hour, top 200



Note: The graph depicts the average standardized *individual skill relative to opponents* in the groups, including 95% confidence intervals. The top row depicts groups identified on 5% level, the bottom row those identified on 1% level. The parameters of the analysis vary across the different figures (i.e., bracketing of 1 hour versus 2 hours as well as the top 100 most frequent players versus the top 200 most frequent players of each version).

Table 19: Performance regression - success when behind

	Chasing		Unbiased	
	won	won	won	won
behind_before_match	-0.00870**	-0.00642*	0.00093	0.00216
	(0.00369)	(0.00341)	(0.00242)	(0.00238)
match_in_session		-0.00005		-0.00012
		(0.00005)		(0.00008)
high stakes		-0.00508		-0.00248
		(0.00479)		(0.00310)
skill-dependent environment		0.03561***		0.02256***
		(0.00717)		(0.00526)
_cons	0.53338***	0.52745***	0.52192***	0.52156***
	(0.00385)	(0.00416)	(0.00182)	(0.00263)
N	91170	91170	203761	203761
adj. R^2	0.000	0.001	-0.000	0.000

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table provides the results of regression specification (1) and session bracketing of one hour. It considers individuals classified on 5% level.

Table 20: Performance regression - success when ahead

	House money		Unbiased	
	won	won	won	won
ahead_before_match	0.00428	0.00619	-0.00066	0.00015
	(0.00592)	(0.00602)	(0.00246)	(0.00250)
match_in_session		-0.00025		-0.00012
		(0.00018)		(0.00008)
high stakes		-0.02341**		-0.00246
		(0.01157)		(0.00308)
skill-dependent environment		0.02034*		0.02238***
		(0.01120)		(0.00524)
_cons	0.50445***	0.50781***	0.52261***	0.52243***
	(0.00524)	(0.00581)	(0.00194)	(0.00269)
N	27668	27668	203761	203761
adj. R^2	-0.000	0.001	-0.000	0.000

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table provides the results of regression specification (2) and session bracketing of one hour. It considers individuals classified on 5% level.

Table 21: Regression - time until next session

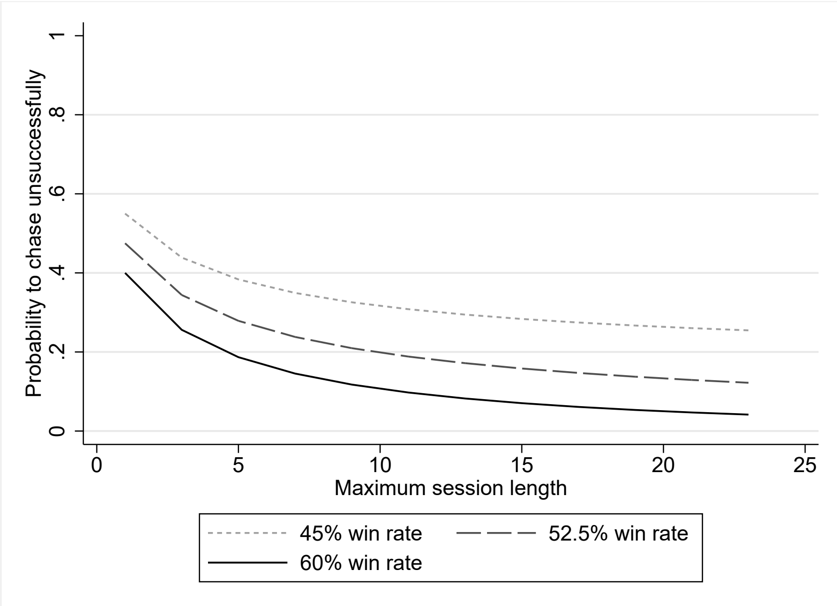
	House money		Unbiased		Chasing	
	time	time	time	time	time	time
lost session	12.39***	12.01***	5.74***	5.91***	1.95	2.07
	(2.31)	(2.30)	(0.96)	(0.96)	(1.33)	(1.31)
match in session		-0.01		0.08		0.05
		(0.19)		(0.08)		(0.08)
high stakes		22.01***		4.72**		0.12
		(5.33)		(1.85)		(2.51)
skill-dependent environment		21.76***		17.74***		3.47
		(3.48)		(2.46)		(2.75)
_cons	21.26***	10.43***	21.22***	15.44***	17.85***	16.29***
	(1.95)	(2.49)	(0.90)	(1.19)	(1.14)	(1.57)
N	5819	5819	28688	28688	11164	11164
adj. R^2	0.006	0.026	0.002	0.009	0.000	0.001

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

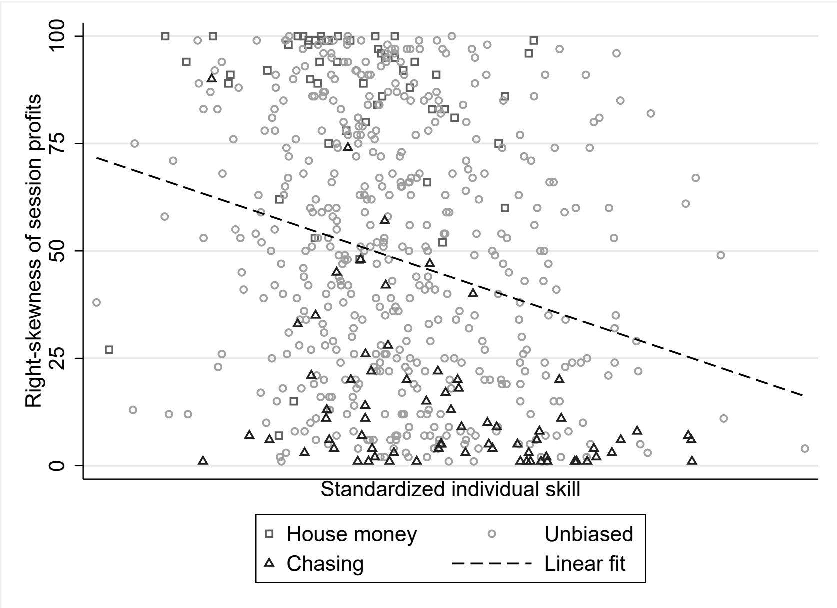
Note: The table provides the results of regression specification (3) and session bracketing of one hour. It considers individuals classified on 5% level.

Figure 9: Probability to chase unsuccessfully depending on maximum session length by individual success rate



Note: The graph depicts the relationship between maximum session length and the probability to chase unsuccessfully depending on different levels of individual success rates, i.e. the probability to win a match $p_w \in \{0.45, 0.525, 0.6\}$. These values correspond to the winning probabilities of the 5%, 50% and 95% percentile players in the observed data. The results are based on simulations based on the assumption that individuals play until they won one more match than they lost in the session (a successful chase) or until they reach the maximum session length without reaching this outcome (an unsuccessful chase).

Figure 10: Relation between *individual skill relative to opponents* and right-skewness of session profits



Note: The graph depicts the relation for the 200 most frequent players of each poker version. Individual skill relative to opponents is standardized with respect to the different versions. Right-skewness of session profits is measured using simulations. In particular, session lengths and win rate are held constant for each individual, while the matches are randomly shuffled 99 times. I report the rank (percentile) of the observed skewness of session profits when compared to those of the simulation and refer to it as right-skewness of session profits. The negative relation of right-skewness of session profits and standardized individual skill is statistically significant ($p = 0.012$). To facilitate comparison, the graph employs different symbols depending on the classification of players.