EMPIRICAL ARTICLE

A longitudinal analysis of the hot hand and gambler's fallacy biases

Brian A. Polin [ID] and Eyal Benisaac

Jerusalem College of Technology, Jerusalem, Israel

Corresponding author: Brian A. Polin; Email: polin@jct.ac.il

Abstract

Researchers have found evidence of both hot hand and gambler's fallacy biases in lottery number selection. Which of the two opposite effects is observed is often dependent upon the nature of the lottery game, the particular sample, the local culture of the participants, or the time transpired since the seed event. By observing hundreds of millions of lottery entries over 118 consecutive semiweekly drawings, we present evidence of both effects and their longitudinal properties. With respect to the selection of *individual* numbers, lottery participants tend to avoid recently selected winning numbers. This gambler's fallacy effect diminishes and the number becomes increasingly 'hot' until it is selected again. With respect to winning number *combinations*, we found strong evidence of a small but persistent hot hand bias. This bias gradually diminishes over time, but remains detectable and highly consistent for a number of years.

1. Introduction

The notion of independent statistical events is often taught in the first lecture of the introductory level course in probability. If one flips a fair coin and it lands on its head (or tail), the likelihood of any subsequent flip resulting in a head (or tail) remains unchanged. This is because each event is independent, and the likelihood of an outcome is uninfluenced by previous ones. Many forms of gambling, roulette and lotto among them, are independent by design. When betting, gamblers often employ heuristics that violate the principle of independence. A lottery participant employing the *gambler's fallacy* heuristic will avoid selecting numbers that were chosen as winners in prior draws. Conversely, a lottery participant guided by the *hot hand fallacy* will prefer numbers that were chosen as winners in prior draws.

Sports fans, even of the non-gambling variety, are familiar with the term 'hot hand'. A basketball player who successfully makes multiple consecutive baskets is said to have a hot hand. Arguably, Gilovich et al. (1985) were the first to investigate the phenomenon. They examined whether the chances of hitting a shot were greater after a prior successful attempt than a prior failed attempt. As the likelihood of completing a shot in basketball is a function of the skill of the shooter (and the defender), the basketball-derived term does not apply to truly random events. It should be noted that Miller and Sanjurjo (2018) revisited Gilovich et al.'s (1985) research and found a flaw in their methodology, thereby overturning the result, and presenting real evidence of a hot hand effect. Sundali and Croson (2006) distinguish between hot hand and hot outcome, with the latter indicating a

preference for winning numbers from prior lottery draws or roulette results. In keeping with convention though, we use the terms 'gambler's fallacy' and 'hot hand' for the duration of this article, even if there is no actual anatomical hand influencing the outcome of the event.

In their analysis of the Maryland state lottery, Clotfelter and Cook (1993) found evidence of the gambler's fallacy. The lottery game investigated involved the selection of a three-digit number. They discovered a 'clear and consistent tendency for the amount of money bet on a particular number to fall sharply after it is drawn, and then gradually recover to its former level over the course of several months'. The 'return to normal' generally occurred after 84 draws, or 84 days, as the game under investigation involved a daily draw.

Terrell (1994) found similar evidence of the gambler's fallacy in the New Jersey three-digit lottery, although to a lesser extent than Clotfelter and Cook's (1993) findings from the Maryland lottery. Another factor distinguishing between the Maryland and New Jersey games is the size of the payout. Maryland's lottery offers a fixed $250 prize for correctly guessing a three-digit number. New Jersey's game is pari-mutuel, where the payout and the total wagers on a particular number are inversely related. The pari-mutuel nature of New Jersey's lottery may explain why the gambler's fallacy is less pronounced than in Maryland. For gamblers, the 'unattractiveness' ('attractiveness') of a particularly number may be offset by a larger (smaller) payout.

Rather than investigating preference for or avoidance of winning numbers, Guryan and Kearney (2008) examined winning stores. They discovered a significant increase in lottery ticket sales, with respect to average ticket sales, over a 30-month period in Texas stores that sold winning lottery tickets. The 'lucky store effect' correlated positively with the size of the jackpot and economically disadvantaged populations. They found the effect to dissipate over time, but remained detectable for up to 40 weeks. The authors juxtapose this finding with those of Clotfelter and Cook (1993) and Terrell (1994) in their effort to reconcile evidence of the hot hand that seems to contradict earlier evidence of the tendency of lottery participants to employ the gambler's fallacy heuristic. 'When do individuals subscribe to the hot hand fallacy versus the gamblers' fallacy?' (p. 468)

By analyzing payout rates, Kong et al. (2020) observed the gambler's fallacy in a Chinese lottery. In a daily three-digit game, bettors tended to avoid past winning numbers, until they bounced back to their initial levels 'after around 60 draws'. These findings confirm, and largely replicate, those of Clotfelter and Cook (1993). Kong et al.'s (2020) analysis of a four-digit thrice-weekly Chinese lottery generated very different results and provided evidence of the hot hand fallacy. In this game, 'bettors actively seek past winning numbers', which return to their initial levels after 25 draws.

In contrast with the prior studies, Ho et al.'s (2019) investigation was based on the complete data set from the 203 drawings of Taiwan's twice weekly C(42,6)[1] lotto game held in 2002–2003. They too found evidence of the gambler's fallacy, with winning numbers underrepresented among guesses in subsequent draws. Their rich data set also enabled them to uncover more nuanced phenomena. The intensity of the gambler's fallacy varied with the frequency of a number's past winning history. For numbers chosen 'frequently' in the past as winners, the gambler's fallacy was less intense than for numbers chosen 'infrequently' in the past as winners.

While the prior mentioned research, all lottery-based, found evidence of a singular form of selection bias at the aggregate level, Sundali and Croson (2006) found evidence of both contrasting biases at the personal level. Aggregate-level lotto research determines whether a number is over- or under-represented with respect to its global popularity after it has been selected as a winner. It is likely that some individuals prefer winning numbers and others avoid them, but most lotto research is unable to untangle the two abrogating biases, and reports only on the aggregate deviation, positive or negative, from a number's global mean over a defined period. Exceptionally, Dillon and Lybbert (2021) identified both biases, with 6.3% and 15.7% of lottery participants in Haiti and Denmark, respectively, preferring recent winning numbers, but 'average' players in both countries avoiding recent winning numbers.

---

[1]This is standard combinatoric notation and should be read '42 choose 6', or select six numbers from a table consisting of numbers one to 42. This particular Taiwanese game consists of 5.2 million possible combinations.

In their casino roulette-based research, Sundali and Croson (2006) found evidence of a tendency among some individuals to prefer past winning numbers, while other individuals consciously avoided them. Both groups constituted 50% each of the sampled population. In earlier roulette-based research, Croson and Sundali (2005) found evidence of the gambler's fallacy. After streaks of identical outcomes, gamblers tend to bet (disproportionally) against a recurrence of that outcome. After experiencing a win, gamblers tend to 'bet on more numbers'. This finding addresses the sum of money bet, but not a preference for betting on the prior winning numbers.

Similar to Croson and Sundali's (2005) discovery in an actual casino, Ayton and Fischer (2004) found evidence of both positive and negative recency effects in a simulated casino experiment with university students. Subjects demonstrated the gambler's fallacy by betting against a recurrence of an outcome, but with multiple recurrences of this outcome, they tended to bet in favor of the continuation of the streak. Suetens et al. (2016), based on the Danish lotto's C(36,7) weekly game, also found evidence of the gambler's fallacy morphing into the hot hand, with a number's increasing 'hotness'. Players bet 1.6-3% less on numbers drawn in the previous week's lottery, but with each subsequent re-draw, their popularity increased by 0.9-1.4%. Wang et al. (2016) demonstrated an avoidance of recently drawn numbers in the Dutch lotto. Their analysis of hot numbers revealed hot hand tendencies among infrequent players and simultaneous gambler's fallacy behavior among frequent players.

In an effort to reconcile the seeming contradiction between the evidence of simultaneous (as opposed to sequential) hot hand and gambler's fallacies, Ji et al. (2015) considered cultural influences. In a coin-tossing experiment at a Canadian university, Asian students demonstrated the gambler's fallacy, expressing the belief that their luck would turn after three incorrect predictions. Caucasian students behaved in accordance with neither the hot hand nor the gambler's fallacy biases. In a second experiment involving consecutive basketball hits or misses, Caucasians were more likely than Asians to demonstrate the hot hand bias and predict the continuation of the streak. The aforementioned research is summarized in Table 1.

By analyzing a large body of archival data consisting of hundreds of millions of actual gambler preferences, we hope to provide further nuance to the discussion on recency biases. The nature and the volume of the data, to be detailed in the next section, offer an opportunity to identify subtle phenomena of a longitudinal nature. Specifically, we seek to identify the presence of a bias and quantify its parameters, including the direction of the bias, positive or negative, the duration of the bias, and its rate of abatement.

2. Data

2.1. Background

To address these questions, we analyzed manually selected numbers from the 118 'Lotto' drawings in 2018. Lotto is the flagship game of the officially sanctioned Israeli National Lottery, Mifal HaPayis (MhP). Nationally televised drawings are held on Tuesday and Saturday nights, with additional drawings for national holidays and other special occasions. The data set upon which this research is based may not be disclosed, as the lottery authority conditioned the release of the data to us upon signing a nondisclosure agreement. Upon request, aggregated data may be made available to researchers, if permission is granted by the legal department of the MhP.

Lotto participants select six non-repeating numbers ranging from 1 to 37, and an additional power number ranging from 1 to 7. The grand prize is awarded when all numbers (6+1) match the balls drawn on the live television broadcast. Smaller prizes are awarded for correctly guessing 3 or more of the winning numbers. See the truncated sample form presented in Figure 1.

According to MhP officials, manual guesses, numbers chosen individually by Lotto participants, constitute approximately 50% of all guesses. 'Lottomat' is also available, whereby participants check a box on the paper form indicating a preference for computer-generated numbers. Participants hand the form to the Lotto booth attendant, who collects payment and scans the form into the system. Payment

Table 1. *Summary table of past recency bias research.*

	References	Game	Location	Summary of findings
1	Clotfelter and Cook (1993)	Lottery	Maryland	3-digit game, negative bias, returns to normal after 84 days
2	Terrell (1994)	Lottery	New Jersey	3-digit game, negative bias, pari-mutuel, bias remains for 60 days
3	Guryan and Kearney (2008)	Lottery	Texas	Positive preference for stores where winning lottery tickets were sold, duration 40 weeks
4	Kong et al. (2020)	Lottery	China	Hot hand occurs in multi-prize games, gambler's fallacy in single prize games. Gambler's fallacy remains for 60 draws in 3-digit game, hot hand lasts 25 draws in 4-digit game
5	Ho et al. (2019)	Lottery	Taiwan	Complete database of 6-digit game. Gambler's fallacy observed, magnitude of effect varies based on a number's popularity
6	Sundali and Croson (2006)	Roulette	Nevada	A sample of 136 gamblers, some hot hand, some gambler's fallacy, in equal proportions. Most of the sample population demonstrated no significant recency bias
7	Dillon and Lybbert (2021)	Lottery	Haiti, Denmark	In both countries, some gamblers demonstrate hot hand, but gambler's fallacy is a more dominant tendency. Individual-level data shows two distinct types of players
8	Croson and Sundali (2005)	Roulette	Nevada	Gambler's fallacy, increases with a prolonged streak of 5 consecutive outcomes of a particular outcome
9	Ayton and Fischer (2004)	Roulette, random events	London	Students demonstrated gambler's fallacy in their predictions of future binary roulette outcomes and hot hand in prediction of random events
10	Suetens et al. (2016)	Lottery	Denmark	7-digit game, individual-level data. Players bet less on past winning numbers (gambler's fallacy), but more on the same numbers (hot hand) as the numbers repeat and become hotter
11	Wang et al. (2016)	Lottery	Netherlands	Past winning numbers chosen less often (gambler's fallacy), repeatedly chosen numbers are preferred (hot hand). Infrequent players behave according to hot hand, frequent players behave according to gambler's fallacy
12	Ji et al. (2015)	Coin toss, basket-ball	Canada	Asian students more likely than Euro-Canadians to use gambler's fallacy in predicting outcome of coin toss. Euro-Canadians more susceptible than Asians to hot hand fallacy in predicting the outcome of shots in basketball

is based upon the number of completed tables, with each table on the form representing an independent guess. A form may contain anywhere from 2 to 14 completed tables. Players preferring lottomat must indicate a preference for 10, 12, or 14 completed tables or guesses. Each completed table costs 3 NIS (1 NIS = $0.30 US). A fully completed form consisting of 14 submitted guesses costs 42 NIS. In 2018, the minimum advertised grand prize was 4 million NIS. With rollover, the grand prize reached
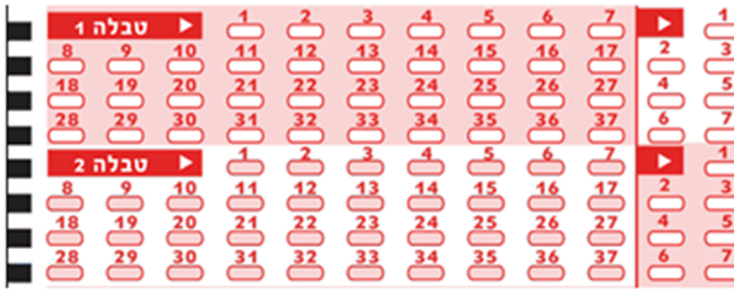
Figure 1. *Sample Lotto form (for brevity, only 2 of 14 tables are shown).*

Table 2. *Data set summary information for the 118 drawings held in 2018.*

	Max	Min	Mean	SD
Advertised grand prize	28,000,000 NIS	4,000,000 NIS	12,398,305 NIS	7,329,999 NIS
Number of (manual) guesses	1,974,046	474,507	973,156	283,088

a maximum of 28 million NIS on a number of occasions. The grand prize was awarded, in some cases to multiple winners who shared it, in only 19 (~16%) of the 118 drawings. The number of manual guesses per drawing ranged from a low of half a million to a high of nearly two million, for a total of 115 million guesses over the course of the year, with each guess consisting of the selection of six individual numbers and an additional power number. This constituted over 800 million conscious selections, and enabled the detection of subtle but significant selection biases. The number of guesses submitted correlated very highly with the size of the advertised grand prize. As MhP was unwilling to share unique form numbers, we were unable to determine the actual number of *people* or forms behind the guesses. It is likely that a greater number of guesses per draw is a reflection of both more guesses per participant and more unique participants. See Polin et al. (2021) for a complete discussion on this linkage. Table 2 features summary statistics of guesses, drawings, and prizes.

To determine the impact of winner status on a number's popularity in subsequent draws, it was necessary to first establish a baseline popularity for each of the 37 numbers in the table. Not surprisingly, and consistent with the extant literature, number popularity was far from uniform. The number 7 was consistently the most frequently preferred number among lotto participants submitting manual guesses. This was the case in every one of the 118 drawings. Conversely, 37 was the most unpopular number in every one of the drawings. An 'averagely popular' number would be chosen (1/37=) 2.7% of the time.

On average, the number 7 constituted more than 3.5% of the numbers selected, while 37 constituted less than 2%. Figure 2 presents the maximum and minimum frequencies for each of the 37 numbers over the course of the 118 drawings in 2018. As may be seen from the figure by comparing the 'max' and 'min' columns for each number, the relative popularity of each number is very consistent and robust. All popularity fluctuations are between 0.10% and 0.23%.

In addition to preference for, or avoidance of, particular numbers, there was a consistent row effect. The average popularity of numbers in the first row of the form (1–7) was 3.13%. This was followed by 2.95% (numbers 8–17), 2.68% (numbers 18–27), and 2.10% (numbers 28–37) for numbers in the second, third, and fourth rows of the form, respectively. The relative popularity of the rows effect was consistent and universal thought every one of the 118 draws in the data set.

2.2. Individual numbers

We next determined the popularity for each of the 37 numbers in each of the 118 drawings, relative to the number's global (or average) popularity. For a given draw, a number selected more often than its
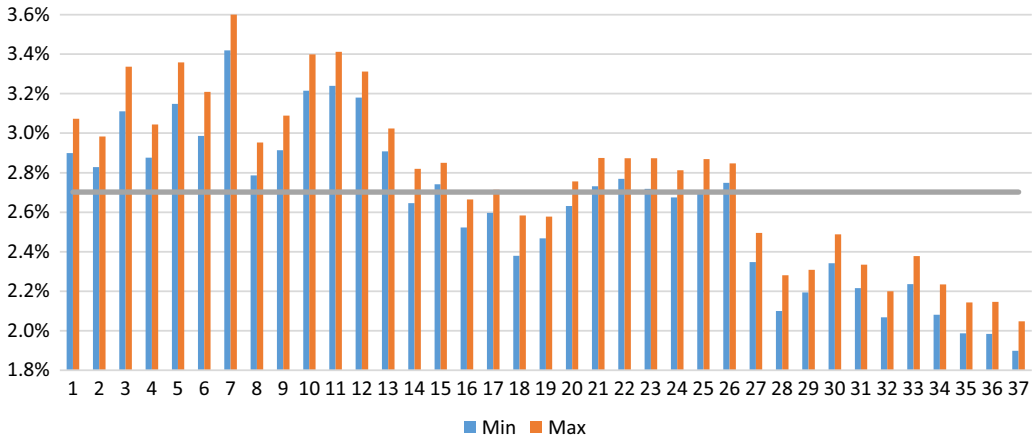
Figure 2. *Maximum and minimum frequencies of numbers manually selected by Lotto participants over 118 draws.*
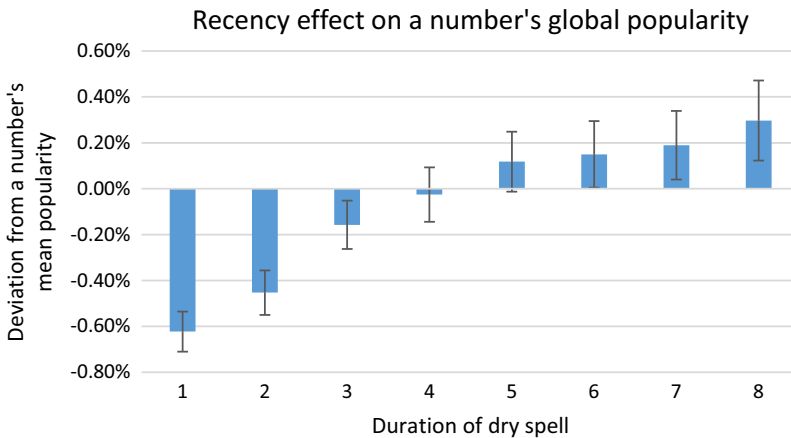

Figure 3. *Deviations from global number popularity as a function of dry spell with 95% confidence intervals.*

global frequency is positive, while a number selected less often than its global frequency is negative. For lottery participants who are interested, past winning numbers are easily accessible to the public on the MhP website. Although the size of the grand prize is widely advertised in order to encourage participation, past winning numbers, though publicly available, are *not* advertised. For recency biases to occur, participants must be aware of prior winning numbers, and must actively seek them online, or track and record them as they are reported in the media after each drawing. The percent deviation from the global popularity was recorded for every one of these ($37 \times 118 =$) 4,366 observations. These values constitute the y-axis values in Figure 3. Each one of these observations was paired with an x-value, the number of drawings since the number (1–37) was last chosen as a winner. The longest gap, or dry spell, between repeat selection of a particular number as a winner was 51 consecutive draws. In this particular case, one of the winning numbers in drawing number 3013 held on the 8th of May 2018 was 14. The previous time the number 14 was drawn as a winner was on the 28th of November 2017.

In general, as the duration of the dry spell increases, the number of observations becomes more sporadic. With so few observations, and with such random variability in number popularity, the mathematical function describing this phenomenon may have very little reliability or predictive value.

To overcome this problem, it was necessary to determine a truncation point for a maximum dry spell after which the observations were disregarded.

The maximum dry spell was recorded for each of the 37 numbers on the Lotto form. Every one of the 37 numbers experiences a dry spell of at least 12 consecutive drawings. This was the first truncation point in determining the gap duration for analysis, as including analysis of gaps longer would be only a partial representation of the full range of numbers 1-37.

The minimum number of observations was next recorded for numbers 1–37 for each of the dry spell durations ranging from 1 to 12. In order to avoid wild fluctuations in a variable system, we determined that at least three observations of every one of the numbers 1–37 are necessary for each of the dry spell durations. Thus, given our data set, the period of analysis was truncated at 8. No further analysis in the article regarding the decay of the gambler's fallacy exceeds this time horizon for individual numbers. Number combinations, the array of six numbers chosen from the table of numbers 1–37, will be addressed separately later in this article.

The bar at the extreme left of Figure 3 indicates the reduced popularity for numbers 1–37, with respect to their global popularity over 118 draws, in the draw immediately after a number was selected as a winner (gap = 1). The reduced number popularity gradually decreases, until the global popularity is exceeded after four or five draws.

In addition to their early documentation of the gambler's fallacy, Clotfelter and Cook (1993) investigated whether the intensity of the effect varied based on a number's general popularity. As the game they analyzed was a 3-digit game, the number of distinct guesses was 999. Each of these numbers was classified as more or less popular than the mean. Without testing for statistical significance, they found the gambler's fallacy to be stronger among popular numbers. As in their case, number popularity in our data sample varied greatly. As presented in Figure 2, number 7 was chosen most frequently, on average 3.51%. The number 37 was the least frequently chosen number, at 1.96%. Despite strong evidence of the gambler's fallacy, and large gaps in number popularity, we found no evidence of differing intensities recency biases with respect to the popularity of the number.

2.3. Number combinations

With respect to individual numbers, lotto participants choose from among 37 alternatives. When it comes to combinations, lotto participants select from among [C(37,6)=] 2,324,784 alternatives. Given an average of slightly less than one million manual guesses per draw (see Table 1), and more than two million number combination alternatives, assuming a uniform distribution, the likelihood of a random number combination being selected by a participant in a given draw is slightly less than 0.5, or more precisely (973,156/2,324,784=) 0.42[2] . In reality though, just as preferences for individual numbers are not uniformly distributed, participants demonstrate a preference for certain combinations. These combinations are often consecutive numbers, uniform gaps between numbers, combinations that form diagonal lines, or v-shapes based on the dimensions of the lotto form (see Figure 1). A figurative histogram would consist of 2.3 million bins, most of which would be completely empty, while at the other extreme, a small percentage of the other bins would contain dozens of observations. In our data set, the three most popular number combinations, each chosen cumulatively more than 5,000 times over the course of 118 drawings were {1,2,3,4,5,6}, {32,33,34,35,36,37}, and {5,10,15,20,25,30}. A full discussion on biases in the selection of number combinations is beyond the scope of this article, but is discussed extensively in the extant literature (Baker and McHale, 2011; Becser and Zoltay-Paprika, 2016; Farrell et al., 2000; Hauser-Rethaller and König, 2002; Henze, 1997; Krawczyk and Rachubik, 2019; Papachristou and Karamanis, 1998; Wang et al., 2016).

To detect the presence of a number combination selection bias, we tracked the popularity of the winning combinations from the 339 draws held between January 2016 and December 2018 in

---

[2]When including the 'power number', a smaller table consisting of numbers 1–7, the actual number of possible combinations increases from 2.3 million to 16.3 million. A full discussion on the selection biases of power numbers is beyond the scope of the paper, but is discussed in greater detail in Polin et al. (2021).
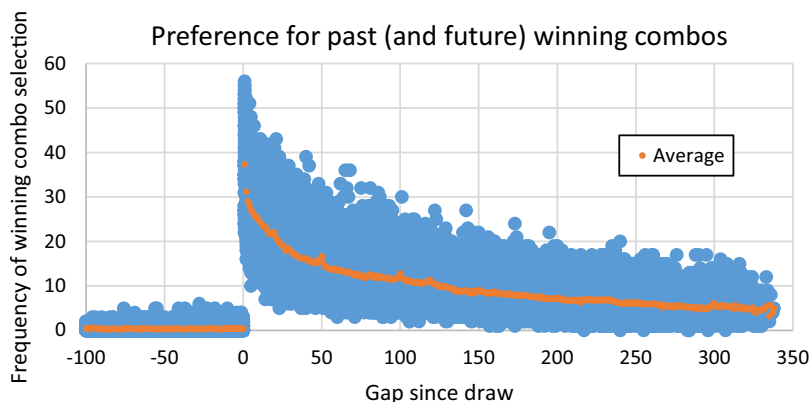
Figure 4. *Preference for winning number combos as a function of draws (pre- and) post selection.*

the manual selection of lotto participants in the 118 drawings held in 2018. The duration of the gaps between a drawn winning combination and its selection by participants ranged from minus 117 draws (the winning combination from the last draw of 2018 and its selection frequency by lotto participants in the first draw of 2018) to 338 draws (the winning combination from the first draw of 2016 and its selection frequency by lotto participants in the last draw of 2018). A negative gap implies a player's selection of a winning combination *before* it is actually selected as a winner. Obviously, there can be no recency-based selection bias among number combinations that will be selected in the future as winners. The purpose of observing the negative gaps is to provide insight into the frequency of selection of 'arbitrary' number combinations and determine a baseline.

The left-most portion of Figure 4 indicates that preference for future winning number combinations, which can be deemed arbitrary combinations, hovers between zero and five. Once, however, a combination is selected as a winner, it is initially preferred by tens of lotto participants. This is indicated on the portion of the figure to the right of the zero point on the horizontal axis. Ho et al. (2019) also present empirical evidence of a preference for combinations selected as winners in recent past draws.

There is interesting anecdotal evidence of this phenomenon both in the Israeli lottery and other lotteries throughout the world. Hand (2014) presents the case of the same set of winning numbers being chosen in the C(49, 6) Bulgarian lottery on 6 September and again in the subsequent draw on 10 September, 2009. There were no winners in the 6 September drawing, but 18 participants guessed correctly in the 10 September drawing. Due to the pari-mutuel nature of the game, each of the winners was awarded 10,164 levs (~$7,700 US). Similarly, the winning numbers drawn in Israel's lotto on 21 September 2010 were drawn again on 16 October, four weeks or seven draws later. According to the MhP archives, 92 players matched all six numbers and received (only) 4,000 NIS each. An additional three players also matched the power number and each won 4 million NIS. As in the case of the Bulgarian lottery, zero participants guessed this combination of numbers the first time it was drawn. Stefanski (2008) presents the case of North Carolina's Cash 5 winning numbers from 9 July 2007 repeating themselves on 11 July, although the number of participants selecting those 'lucky' numbers is unclear.

According to our data, the preference for past winning numbers gradually decreases over time, but is still observable after 338 draws, or three full years later. In total, Figure 4 contains over 40,000 data points, with the frequency of selection of winning combinations from 339 draws (in 2016–2018) among participants in 118 draws (in 2018). The orange data points represent the average number of times a past winning combination is a manually selected guess for each gap duration. In a descending ranking of combination popularity, with {1,2,3,4,5,6} in first place, over 118 drawings, the winning combination from the previous draw would be in 16th place, with combinations from 2, 3, 4, and 5 draws ago in 43, 51, 54, and 57 places, respectively.

Table 3. *Regression model with mean popularity of past winning combinations as the dependent variable.*

Variables	Model 1			Model 2		
	B	β	SE	B	β	SE
Constant	18.903		.051	12.096		.092
Periods post draw	−.056	−.709	.000	−.057	−.714	.000
Total manual guesses				7.038E-6	.300	.000
R^2	.503			.593		
ΔR^2				.090		

A nonlinear regression trend line fit to the positive data points (those to the right of the zero value on the horizontal access) y = -5.593ln(x) + 36.904 has an R^2 value of 0.656. At the rate of decay predicted by the trend line, the last vestiges of the hot hand disappear after six full years. After six years, the average manual selection frequency of past winning combinations returns to (973,156/2,324,784=) 0.42. Although the preference for previous winning combinations is best described by a decreasing nonlinear function, a linear regression analysis yields interesting results. Model 1, consisting of 'periods post draw' as the lone predictor variable, significantly accounts for >50% of the variance in the times a prior winning number combination is selected. Model 2, including both 'periods post draw' and 'total number of manual guesses' together, significantly accounts for >59% of the variance (Table 3).

3. Discussion

Although the recency bias with respect to the conscious selection of individual lottery numbers is small percentage wise, it is consistent and statistically significant. After a number has been drawn, it is avoided by a small percent of the lottery participants. This gambler's fallacy bias gradually and consistently diminishes over the four subsequent draws. At the rate of two draws per week, after two weeks, 'dry' numbers are no longer avoided. From a dry spell of five, the avoidance of a number once selected as a winner seems to become a preference for the number. This preference increases until the number is drawn again. The trend presented in Figure 3 is an increasing linear function, but it is merely a section within a larger repeating pattern, as every number will inevitably be drawn again. This repeating pattern of a number's decreasing unpopularity followed by its increased popularity, then truncated by being selected again as a winner might be depicted by a saw tooth wave pattern. Indeed a number's popularity increases as the dry spell increases, but it is unclear if the gambler's fallacy morphs into a hot hand bias as suggested by Suetens et al. (2016). It is possible that these are not two successive and opposite recency biases, but rather one continuous phenomenon of varying degrees of 'hotness'. The selection of number facilitates a gambler's fallacy that cools the number. Over time, the number heats up, until 'ambient' temperature is reached, or the number is selected again and the increasing 'hotness' begins anew.

While the lottery form contains only 37 individual numbers, the selection of any six of those numbers allows for 2.3 million number combinations. Contrary to the 37 individual numbers, lottery participants demonstrate a clear hot hand bias and actively prefer strings of numbers recently selected as winners. Although the preference is measurable and significant, in absolute terms, it is even more rare than the recency bias in the selection of individual numbers. It is also less pronounced than a general preference for aesthetic sequences of sequential numbers of fixed gaps. A string of numbers may be preferred by 30–40 or even 60 lottery participants in the draw after it is selected as a winning combination. The popularity of past winning combinations gradually decreases, but they remain more popular than 'average' non-winning combinations years later. In both cases above, individual numbers and number

combinations, we identified statistically significant recency biases and defined the parameters, namely their frequency and their duration.

This paper is not intended to be a 'how to' guide for the novice gambler, but one might indeed use its analysis to increase the expected return on an investment in a lottery ticket, or any other pari-mutuel game. More generally, though, a nuanced understanding of human behavior in decision-making with respect to prior successes or failures may be beneficial in investments, pricing, inventory control, or any other system in which the predominance of a public choice affects the expected return.

Through experimentation, Navarrete and Santamaría (2012) demonstrate a reduced incidence of gamblers' fallacy with an increase in the number of alternatives from which to choose. Similarly, Kong et al. (2020) observed a strong presence of the gamblers' fallacy in 'single-prize' games, but the effect disappeared as the number of outcomes was increased. This, in turn, led to a significant increase in hot hand behavior, as game complexity increased. We have presented empirical evidence of both biases when choosing from among dozens of alternatives and from among millions of alternatives.

3.1. *Limitations and future directions*

As is often the case with experimental or observational research, our findings are a reflection of the data sample. While the sample is large enough to present stable and robust results that would unlikely change with a larger data set, similar research in other countries or cultures may yield different findings.

Another limitation of this research may have to do with the probabilistic nature of the phenomena observed. For events that are very rare, it is difficult to observe changes in their absence. If the expected likelihood of a given C(37,6) number sequence being chosen by at least one player in a given lottery is under 50%, then the non-selection of the sequence may be attributable to 'chance', rather than conscious avoidance. This means that a positive selection bias is easily detectable, while a negative selection bias may be difficult or impossible to detect. This is analogous to drivers swerving to avoid an object on the road. A driver in the center lane may swerve to the left or right. A driver in an outermost lane may swerve only in one direction, even if his 'preferred' direction is the opposite one. This may explain why the dominant selection bias (positive or negative) is often related to the total number of choices one has. This may explain the different and opposite biases overserved in 3-digit versus 4-digit lottery games (see Kong et al., 2020).

Most lottery-based research is at the aggregate level. The detection of a bias in a particular direction may not be a true indication of the intensity of the bias, but rather the net result of the greater of two offsetting biases. Some notable exceptions to population-based lottery research are Otekunrin et al. (2021) and Suetens et al. (2016), where number preferences at the individual level were observed via unique customer ID numbers or in-person interaction between data collectors and bettors at lottery booths. Unfortunately, Israel's MhP does not collect individual-level information. The location of sale, however, is recorded, and each one of the country's 2,400 officially sanctioned points of sale had a unique ID. Despite the generosity of MhP officials in providing us with data, they have not yet made the location of the sale accessible to us. The Israeli Central Bureau of Statistics (CBS) has high-resolution demographic and economic data, and future research investigating number selection biases based on point-of-sale location may yield novel insights. Similarly, controlled experimental research may uncover individual differences and offer insights as to traits and characteristics of gambler's fallacy players versus their hot hand counterparts.

References

Ayton, P. & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition*, *32*, 1369–1378.

Baker, R. & McHale, I. G. (2011). Investigating the behavioural characteristics of lottery players by using a combination preference model for conscious selection. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(4), 1071–1086.

Becser, N. & Zoltay-Paprika, Z. (2016). Patterns in the lottery game. *Forum Scientiae Oeconomia*, *4*(1), 55–70.

Clotfelter, C. & Cook, P. (1993). Notes: The "gambler's fallacy" in lottery play. *Management Science*, *39*(12), 1521–1525.

Croson, R. & Sundali, J. (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty*, *30*, 195–209.

Dillon, B. & Lybbert, T. J. (2021). The gambler's fallacy prevails in lottery play.

Farrell, L., Hartley, R., Lanot, G., & Walker, I. (2000). The demand for lotto: The role of conscious selection. *Journal of Business & Economic Statistics*, *18*(2), 228–241.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*(3), 295–314.

Guryan, J. & Kearney, M. (2008). Gambling at lucky stores: Empirical evidence from state lottery sales. *American Economic Review*, *98*(1), 458–473. https://doi.org/10.1257/aer.98.1.458

Hand, D. J. (2014). Never say never. *Scientific American*, *310*(2), 72–75.

Hauser-Rethaller, U. & König, U. (2002). Parimutuel lotteries: Gamblers' behavior and the demand for tickets. *German Economic Review*, *3*(2), 223–245.

Henze, N. (1997). A statistical and probabilistic analysis of popular lottery tickets. *Statistica Neerlandica*, *51*(2), 155–163.

Ho, H-C., Lee, S-C., & Lin, H. (2019). Modelling of how lotto players select their number combinations dynamically. *International Gambling Studies*, *19*(2), 200–219. https://doi.org/10.1080/14459795.2018.1529814

Ji, L-J., McGeorge, K., Li, Y., Lee, A., & Zhang, Z. (2015). Culture and gambling fallacies. *SpringerPlus*, *4*, 510.

Kong, Q., Granic, G., Lambert, N., & Teo, C. (2020). Judgment error in lottery play: When the hot hand meets the gambler's fallacy. *Management Science*, *66*(2), 844–862.

Krawczyk, M. W., & Rachubik, J. (2019). The representativeness heuristic and the choice of lottery tickets: A field experiment. *Judgment and Decision Making*, *14*(1), 51–57.

Miller, J. B. & Sanjurjo, A. (2018). Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica*, *86*(6), 2019–2047.

Navarrete, G. & Santamaría, C. (2012). Adding possibilities can reduce the gambler's fallacy: A naïve-probability paradox. *Journal of Cognitive Psychology*, *24*(3), 306–312.

Otekunrin, O. A., Folorunso, A. G., & Alawode, K. O. (2021). Number preferences in selected Nigerian lottery games. *Judgment and Decision Making*, *16*(4), 1060–1072.

Papachristou, G. & Karamanis, D. (1998). Investigating efficiency in betting markets: Evidence from the Greek 6/49 Lotto. *Journal of Banking & Finance*, *22*(12), 1597–1615.

Polin, B. A., Isaac, E. B., & Aharon, I. (2021). Patterns in manually selected numbers in the Israeli lottery. *Judgment and Decision Making*, *16*(4), 1039–1059.

Stefanski, L. A. (2008). The North Carolina Lottery Coincidence. *The American Statistician*, *62*(2), 130–134.

Suetens, S., Galbo-Jørgensen, C., & and Tyran, J. (2016). Predicting lotto numbers: A natural experiment on the gambler's fallacy and the hot-hand fallacy. *Journal of the European Economic Association*, *14*, 584–607.

Sundali, J. & Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgement and Decision Making*, *1*(1), 1–12.

Terrell, D. (1994). A test of the gambler's fallacy: Evidence from pari-mutuel games. *Journal of Risk and Uncertainty*, *8*, 309–317.

Wang, T., Potter van Loon, R., van den Assem, M., & van Dolder, D. (2016). Number preferences in lotteries. *Judgment and Decision Making*, *11*(3), 243–259.