Rational and Irrational Belief in the Hot Hand: Evidence from *Jeopardy!*

Anthony J. Kukavica Sridhar Narayanan

This Version: May 2023*

Abstract

A longstanding question in behavioral economics concerns whether a "hot hand" exists in sports, gambling, and related settings. In this paper, we leverage a comprehensive play-by-play dataset from the game show *Jeopardy!* to show that such an effect indeed exists in players' in-game performances. We also find evidence for contestants' belief in a hot hand as reflected in their wagering decisions during gameplay. We find that players overestimate the magnitude of the true effect by approximately 3 to 8 times. We also find lower levels of hot hand bias in more successful players as well as players from more quantitatively demanding professions. Lastly we discuss potential underlying mechanisms that may generate the observed effects.

*[WORKING PAPER – DO NOT CIRCULATE] Kukavica (kukavica@stanford.edu) is a doctoral student in Marketing and Narayanan (sridhar.narayanan@stanford.edu) is a Professor of Marketing and the Younger Family Faculty Fellow for 2022-23, both at the Stanford Graduate School of Business. The authors thank Kwabena Donkor, Samuel Goldberg, Yewon Kim, James Lattin, Ivan Li, Peter Reiss, Navdeep Sahni, Andrey Simonov, Jeffrey Zwiebel, and seminar participants at Stanford University for comments and suggestions that have greatly improved this paper. Kukavica acknowledges generous financial support from the Sakurako and William Fisher Family PhD Research Fellowship, and Narayanan would like to thank the Younger Family Faculty Fellowship for supporting this research.

1 Introduction

Behavioral scientists, financial economists, and psychologists have debated whether a "hot hand," defined as a higher likelihood of current success given recent success (controlling for trial state and individual ability), exists in sports, gambling, and related settings. Gilovich, Vallone, and Tversky (1985) analyzed NBA basketball data and concluded that no such effect exists, contrasting with widely held beliefs by players and fans alike.[1] More recently, Green and Zwiebel (2018) studied MLB baseball data and found that hot hand effects persist across a variety of statistical categories, suggesting that Gilovich, Vallone, and Tversky (1985)'s null result might derive from endogenous defensive responses in the basketball setting.[2]

In this paper, we use novel play-by-play data spanning over 235,000 clues across 3900 episodes of the American trivia game show *Jeopardy!* to answer three focal questions of this literature. First, does a hot hand effect exist in our setting? Second, do contestants *believe* such an effect exists, and if they do, are their beliefs biased? Third, what underlying mechanisms might generate such an effect?

The answer to the first question has varied in the literature but has gradually converged towards the consensus that a hot hand effect generally exists upon properly controlling for confounds. However, due to limitations in the types of data and the institutional features of the contexts in which hot hand effects have been studied, the latter two questions have not received adequate attention despite being essential to obtaining a comprehensive understanding of the topic. Our context and the high level of detail in the data allow us to answer these questions in ways the prior literature has been unable to.

[1] Camerer (1989) shows that fans and sports bettors believe in a hot hand effect in basketball.

[2] See also Bocskocsky, Ezekowitz, and Stein (2014) and Lantis and Nesson (2021) for analyses of NBA basketball data that implement more robust controls than the seminal Gilovich, Vallone, and Tversky (1985) study.

We first test for a hot hand effect in contestant performance by asking whether contestants are more successful following recent success upon controlling for baseline contestant ability as well as the state of the game (e.g., current score and question difficulty). Indeed, we find a significantly higher likelihood of contestants answering questions correctly given recent success. Critically, we note that our definition of a hot hand does not require there to be a causal effect of prior performance on future performance, but rather a more general form of short-term positive serial correlation. This definition of the hot hand has been implied but not consistently reinforced in the prior literature.[3]

Next, we leverage an institutional feature of the *Jeopardy!* game to test whether contestants manifest belief in a hot hand by betting more when they are "hot" in an attempt to maximize earnings. At random moments during gameplay (known as "Daily Double" questions), contestants choose an amount of money to wager before answering a question; they gain the wager amount if correct and lose it if incorrect. We find that contestants wager significantly more money when they are "hot" compared to when they are not, but that the magnitudes of these increases in wager amounts cannot be rationalized by the increased probability of success when "hot." Concretely, we separate contestant belief in a hot hand into a behaviorally consistent "rational" component and an "irrational" bias estimate; we find that contestants overestimate the rational amount by a multiplicative factor between 3 and 8. However, we show that this bias dissipates for more successful contestants as well as contestants from professions with more quantitatively demanding professions.

Finally, we explore potential mechanisms that might be generating the effects on contestant performance and wagering behavior. Gilovich, Vallone, and Tversky (1985) speculate that increased confidence could provide a reasonable explanation for the pres-

[3]Note that in settings with identical, repeated trials, a hot hand effect can be viewed as a form of short-term positive serial correlation.

ence of hot hand effects. Fatigue, alertness, and other behavioral tendencies such as limited cognition are also plausible candidates. In our setting, we use natural interruptions during gameplay (commercial breaks) to test whether hot hand effects dissipate with time. We find that the observed hot hand effect disappears in the short-term aftermath of commercial breaks, consistent with mechanisms such as alertness but not with confidence. We also find that contestants from more quantitative or analytical professions exhibit lower levels of bias when assessing the magnitude of hot hand effects on performance.

The *Jeopardy!* game provides a uniquely attractive setting to study questions related to hot hand effects. First, it is a field setting with heterogeneous participant characteristics where large sums of money are at stake, resembling what might be expected of stock market investors or sports professionals much more closely than laboratory experiments. Second, the randomized placement and timing of Daily Double clues mitigate concerns for selection in game state. Finally, we note that much of the prior literature examines either sports or carefully controlled experimental domains.[4] Our setting moves away from sports data and into a setting where we can directly observe contestant behavior regarding their own performance. Given our ability to identify both existence of and belief in hot hand effects simultaneously, we are able to provide a comprehensive test of whether belief in a hot hand is rational, biased, or completely fallacious.[5]

The conclusions of earlier studies led to the hot hand phenomenon being dubbed a

[4]Examples of non-experimental domains are Croson and Sundali (2005) and Narayanan and Manchanda (2012), which empirically analyze roulette and slot machine data, respectively. However, in the case of slot machines, any belief in a hot hand is inherently irrational as an underlying effect cannot exist in such games of pure chance. Guryan and Kearney (2008) examine a version of the hot hand effect, studying the impact of a winning lottery ticket sold at a store on subsequent sales of lottery tickets at that store. Again, an underlying "lucky store" effect cannot exist in such a context, and therefore such a belief must be irrational.

[5]Jin (2018) begins examination of this question in the context of competitive darts, though limitations include a non-comprehensive definition of the hot hand and a setting that likely cannot generalize to financial domains.

"fallacy" by both academics and popular media. Our study demonstrates that this is not the case. Indeed, people tend to have inaccurate beliefs about how having a "hot hand" impacts their probability of success. Despite hot hand effects, the magnitudes of the effects do not justify the magnitudes of the increases in observed wager amounts. However, note that contestants are fully justified in their implicit assessment that a hot hand exists, but with the important caveat that there is a substantial overestimation in most circumstances. In turn, our results suggest that it is more accurate to think of the hot hand phenomenon as a *bias* rather than a *fallacy*.

Our paper's main contribution is threefold: first, we confirm that an underlying hot hand effect in contestant performances does exist in our setting and that players believe in such an effect as manifested through their wagering decisions. We then find that contestant beliefs do not align with reality; instead, they systematically wager 3 to 8 times higher than purely rational belief in the hot hand would predict. We further show that the this bias is smaller for more successful contestants, and for contestants with a quantitative background. Finally, we show that a hot hand effect does not exist after regularly spaced breaks in gameplay and that it reappears shortly thereafter, suggesting that the existence of these effects may be time-sensitive in nature as opposed to, for example, a steady accumulation of confidence due to answering many questions correctly. We also note that our general framework of isolating and quantifying hot hand effects can be adapted to related settings such as investor belief in stock performance, managerial predictions of sports outcomes, and wagering tendencies in other games of skill such as poker.

The rest of the paper is structured as follows: Section 2 describes our setting and data; Section 3 discusses definitions of a hot hand effect and how it might manifest; Section 4 provides evidence for objective existence of a hot hand effect as well as contestants' subjective belief in it; Section 5 develops and estimates an empirical framework for measuring the degree of hot hand bias in our setting; Section 6 explores potential mechanisms that might generate a hot hand; Section 7 concludes.

2 Data

2.1 Description of the *Jeopardy!* Game

Jeopardy! is a popular American trivia game show that airs one 30-minute episode every weekday evening on the ABC television network. It is a unique quirk of *Jeopardy!* for clues to be phrased as answers to questions, upon which players must give their responses in the form of a question. For example, a clue might read "the capital of California," upon which the correct response is "What is Sacramento?" From here onward we refer to clues from the game board as "questions" (or occasionally "clues") despite the fact that these are actually answers to unknown questions that the contestants attempt to recover. Analogously, we will refer to the responses the contestants give as "answers," despite their required phrasing as questions.

Each 30-minute episode contains one game with three contestants, one of whom is the winner of the previous game. There is also a show host whose primary role is to read questions aloud. The contestants then attempt to answer the questions correctly. Each game contains three rounds: a first round, a second round known as "Double Jeopardy," and a third round known as "Final Jeopardy." The first round consists of 30 questions distributed across 6 categories, for example "Movie Stars" or "World Capitals." The categories vary from game to game. Within each category, each of the six hidden questions contains a dollar amount that is visible to contestants before the question is revealed. These are mapped 1-to-1 with the set $\{\$200, \$400, \$600, \$800, \$1000\}$, where the question difficulty increases in expectation with the dollar amount. The categories and dollar amounts are displayed on a game board resembling that shown in Figure

EDIBLE RHYME TIME	BOOKS IN GERMAN	3 “T”s	CHOP CHOP!	THEY SAID IT WOULDN’T LAST	THEY WERE RIGHT
$200	$200	$200	$200	$200	$200
$400	$400	$400	$400	$400	$400
$600	$600	$600	$600	$600	$600
$800	$800	$800	$800	$800	$800
$1000	$1000	$1000	$1000	$1000	$1000

Figure 1: Example Game Board Layout

This figure shows an example of the *Jeopardy!* game board for the first round of gameplay. Each column represents a different category of clues. The top row indicates the category names for the round, and each rectangle with a monetary value on top has a clue beneath it. Note that the category names change across rounds and games, and the clue values are doubled during the second round.

1. To begin the game, the returning champion picks a question, choosing any of the questions on the board. The question is then read aloud by the host.

Each contestant holds a buzzer that allows them to indicate their desire to answer a question. The first contestant to press their buzzer once a question is read aloud by the show host attempts to answer the question. If correct, the contestant wins the dollar amount of the question and loses that amount if incorrect. The contestant answering correctly then gets to pick the next question to go to. If nobody attempts to answer a question or all attempts are incorrect, then the contestant that selected it chooses the next one.

In the first round of gameplay, there is one hidden question on the game board denoted as the "Daily Double" that, when revealed, gives the contestant who chose the question the chance to wager any dollar amount up to their running score (or up to $1000, whichever is greater) before seeing the question. One of the questions on the game board is randomly assigned to be the Daily Double, and this is only revealed when a contestant picks that question on their turn. If the contestant answers the question correctly, then the wagered amount is added to the score and if incorrect, it is subtracted from the score.

After another short break, the Double Jeopardy round follows the same format as the first round with a few minor modifications. First, the contestant with the lowest running score at the beginning of the round picks the first question from the game board as opposed to the contestant who most recently answered a question correctly. Secondly, all dollar amounts of questions are doubled. The rules of the Daily Double remain the same except that contestants can wager the higher of their remaining total or $2000. Thirdly, there are two hidden Daily Doubles in this round instead of one. Like in the first round of the game, the Daily Doubles in this round are randomly assigned to questions on the board; whether or not a question is a Daily Double is only revealed after that question is chosen by a contestant and before the text of the question is shown.

The final round consists of a wagering phase followed by a single question. All contestants with a positive score at the end of Double Jeopardy participate and can wager up to the maximum of their running total and $1000. Contestants know the category of the question, but not the question itself, before making wagers. Upon finalizing wagers, the question is read aloud and contestants have thirty seconds to write down an answer. As before, contestants answering correctly gain their wager, while contestants answering incorrectly lose their wager.

Once Final Jeopardy has concluded, the player with the highest final score wins the game and gets to take home the exact dollar amount of their final score. The player with the second-highest final score wins $2000, and the third place finisher takes home $1000. In the case of a tie for first place, the first among the tied players to press their buzzer and correctly answer a tiebreaker question wins the game. Ties for second and third result in the dollar winnings being split amongst those tied. In this paper we do not analyze results from occasional exhibition games and tournaments where the game rules may differ either minimally or substantially from the norm.

For our empirical analysis, we focus on Daily Double clues in the first two rounds. These are the unique moments in the game when a contestant can make a wager of their choosing, there is no immediate competition during the question (e.g. another contestant trying to press the buzzer), and the placement and difficulty of the questions are random. These properties reduce the potential for confounds present in the dynamic portions of the game, thereby allowing us to attribute our results to variation in the key features of interest: wager amounts, question correctness, whether or not a contestant is "hot," and other potentially relevant determinants of strategy. This aspect of the game is novel to our context and is not present in the typical sports contexts where hot hand effects have been studied. These prior contexts allow for the empirical examination of the existence of a hot hand, but since the players are typically not also making any wagers or other financial decisions at that time, simultaneous examination of belief in a hot hand is harder. Alternatively, belief in the hot hand has been studied in gambling contexts such as slot machines and roulette, but the existence of the hot hand is ruled out by construction in these contexts. Our study is therefore one of the first to examine both existence of and belief in a hot hand, allowing for the quantification of a hot hand bias and analysis of the behavioral mechanisms behind these effects.

2.2 Summary Statistics

We analyze data from *Jeopardy!* games aired between September 2004 and June 2022 played under standard rules. The data is considerable in size: it extends across over 3900 games and 235,000 questions. We observe the question type, the order questions are revealed (since players select the questions to open), the question text itself, wager amounts if applicable, and whether a contestant answered the question correctly. We are able to fully recover the progression of each game from the data.

Critically, we focus much of our analysis of player decisions on Daily Double questions, the three random moments during the game when a contestant has the opportunity to wager an amount of their choosing before observing the question, and then either gains or loses that amount upon answering either correctly or incorrectly. Summary statistics for Daily Double questions are provided in Table 1.

Anecdotally, while all *Jeopardy!* contestants are rigorously vetted through an audition process and are highly skilled at trivia, there are still considerable skill and ability gaps between contestants. We take a number of precautions later to account for this when estimating models of contestant behavior, though in theory the inclusion of contestant fixed effects should be sufficient. In our main analysis we use the full data that includes all players. However, for robustness and a smaller range of contestant abilities in the same, we re-run all our analysis on data including only players with one or more lifetime wins in Appendix Section A.3.

We see that the mean score for a player when making a wager is 7301 dollars. The score is relevant because a contestant is allowed to wager up to but not more than their score amount.[6] We see that the mean deficit and lead amounts relative to the leading or second-place players respectively are 4310 and 4210 dollars respectively. The deficit

[6]The only exception to this rule is that a player with a score below 1000 dollars may wager up to 1000 dollars.

Table 1: Summary Statistics for *Jeopardy!* Daily Double Data

Variable	Min.	1st Quart.	Median	Mean	3rd Quart.	Max.
Score	-4200	3000	6000	7301	10400	46816
Lead Amount	0	1000	2600	4210	5600	40314
Deficit Amount	0	1200	3000	4310	6000	37980
Wager	0	1200	2000	2475	3000	25000
Correct Streak	0	1.0	1.0	1.3	2.0	11.0

We provide summary statistics for all Daily Double questions in our data. Scores, lead amounts, deficit amounts, and wagers are reported in dollars. Streak lengths are reported in raw number of consecutive correct question answers.

amount is calculated as the score difference between the contestant answering the Daily Double and the current leader; te lead amount is calculated as the score difference of the first place player relative to current second-place player. The mean Daily Double wager is 2475 dollars, and the range of wagers extends from 0 to 25000 dollars. Finally, we find the mean streak length to be 1.3 questions heading into a Daily Double, where a streak is defined as the number of consecutive questions answered correctly by a contestant. An incorrect answer or non-answer to a question resets streaks to 0.

3 Background

3.1 What is a "Hot Hand?"

The "hot hand" as a general concept was first studied empirically in the context of professional basketball. Casual observers had long anecdotally speculated that certain players experienced streaks of repeated successes (in this case, shots made) at rates not explainable by pure chance. This proposition generated considerable interest, and Gilovich, Vallone, and Tversky (1985) formally tested whether shot patterns deviated from what independent trials would predict. Their conclusion was that save for extreme cases, the belief was unfounded; in other words, a "hot hand" did not exist in basketball.

Gilovich et al.'s definition of a hot hand seems to align with common perceptions of what it is. Nevertheless, the statistical approaches used to arrive at their conclusion have been questioned, most rigorously by Miller and Sanjurjo (2018). More generally, Gilovich et al. simply view the hot hand as first-order serial correlation in trial outcomes. Later studies found that upon introducing suitable controls for game state and a players' long-run abilities, a small hot hand effect on performance does exist.[7]

Our setting, along with many others where a hot hand effect may exist such as investing and sports betting, cannot realistically be modelled via repeated Bernoulli trials. Specifically in *Jeopardy!*, it is rare to have two in-game situations that are identical: lead changes, variation in question difficulty, and other state features of the game show vary the underlying probability of success before considering any potential hot hand effect. However, following Green and Zwiebel (2018) in defining a hot hand as "short-term predictability in performance," i.e. whether recent successes predict a higher likelihood of current success controlling suitably for confounding variables, we stipulate that a hot hand could still exist in such a setting. Thus, a hot hand effect can be viewed as a form of positive serial dependence, but with the important caveat that the form of the dependence may vary over time and that suitable controls be added that capture both external and long-term factors.

Altogether, when discussing the existence of a hot hand, we are interested in the following question: removing confounding factors that might influence the probability of a successful trial, is recent history predictive of current success? Naturally, this necessitates some arbitrary judgment as to what classifies as "recent" versus not. Rather than taking a stance on a particular definition, we test a variety of methods for specifying recency, allowing for robustness and addressing inherent ambiguity in the definition. These include streakiness, streakiness with diminishing effects, and performance over a

[7]See Bocskocsky, Ezekowitz, and Stein (2014) and Lantis and Nesson (2021) who view the hot hand as a controlled form of serial dependence in basketball trials.

recent time window such as the last two, three, or four questions. A hot hand effect, if present, should manifest for most or all specifications in order to be adequately robust. We elaborate on these measures more precisely in Section 3.2.

Related to the question of whether a hot hand exists, a variety of explanations have been suggested for why people might believe in one. Aside from the obvious explanation that a hot hand simply exists, a leading candidate to explain potentially biased beliefs is the Law of Small Numbers, first introduced in Tversky and Kahneman (1971). The theory stipulates that people make decisions about larger samples based on small-sample representations, despite this not necessarily being statistically valid. They hypothesize that such tendencies may be a result of the representativeness heuristic, a shortcut in human decision-making that simplifies judgment about a population by focusing on its salient properties.

Rabin (2002) proposed a model that showed how belief in the Law of Small Numbers can lead to irrational quantitative judgments by a Bayesian agent. He focused on the gambler's fallacy, a close relative of the hot hand where an agent expects the probability of a successful trial to decrease following recent successe.[8] In a follow up paper, Rabin and Vayanos (2010) contended that belief in a hot hand might derive from the gambler's fallacy despite the two seeming to be opposites. They reasoned that if streakiness has occurred to a point where the agent can no longer rationalize the outcomes with their initial beliefs, uncertainty in the underlying probability can be decreased if belief in a hot hand is introduced.

We note that it is difficult to address underlying behavioral mechanisms that drive belief in or existence of a hot hand without first confirming whether neither, one, or both exist. Relatedly, much of the prior literature studying the question of underlying mechanisms as they pertain to a hot hand builds off of the assumption that a hot hand

[8] See Croson and Sundali (2005) and Narayanan and Manchanda (2012) for empirical analyses of casino gambling settings in which people might falsely believe in the gambler's fallacy.

does not actually exist. Providing strong direction towards a resolution of this question is a major motivation for this paper. In the next section, we proceed to describe our setting in detail to elucidate the foundations of our empirical approach.

3.2 How Should We Measure "Hot"-ness?

In order to measure the effect of a contestant being "hot" on either the probability of answering a Daily Double correctly or their wager amount, we first define what constitutes being "hot." Historically the term has remained somewhat nebulously defined in the literature. A common theme amongst these definitions is that they capture a player's recent performance, with the definition of "recent" depending on the setting.[9]

Green and Zwiebel (2018) argue that good identification of a hot hand effect should specifically capture "short-term predictability in performance." In particular, longer measures of performance such as a player's baseline accuracy or entire performance history would not seem to fit conventional views of what constitutes a player being "hot." With this in mind, it makes sense to conceptualize "hot"-ness as a form of short-term serial correlation upon controlling for underlying player ability as well as game state variables.

In our analysis, we test a number of different measures of "hot"-ness. A natural first choice is a contestant's "correct streak" length S_{ijt}, defined as the number of consecutive questions answered correctly by a contestant i in time period t of game j. Under this definition, a longer streak would stipulate that a player is more "hot" than when experiencing a shorter streak. Due to the possibility that correct streaks may exhibit diminishing returns, we also try defining "hot"-ness as $\log(1 + S_{ijt})$; note that we transform streak length by 1 unit due to the possibility that $S_{ijt} = 0$ at the

[9]For example, Green and Zwiebel (2018) use data from a baseball player's 25 most recent plate appearances when constructing most of their measures of a "hot" state.

start of the game or when all players fail to answer a question, leading $\log(S_{ijt})$ to be undefined.[10]

A final definition we test is a contestant's "M-average," defined as the number of correct answers in the last M questions normalized by M. This measure, introduced in Green and Zwiebel (2018), can take on any value between 0–1 and provides greater forgiveness for incorrect answers in that unlike S_t, it does not reset after a single incorrect answer. Instead, it captures a contestant's performance over a time horizon that is neither too small nor too large. The interpretation is that a higher M-average means that a contestant is more "hot" than if they had a lower M-average. In our main analysis we test the three cases $M = 2, 3, 4$.[11]

4 Investigating Hot Hand Belief and Existence

4.1 Existence of a Hot Hand Effect

We first ask whether a hot hand effect exists in contestants' performances on Daily Double clues. We approach this question by testing for a positive effect of being "hot" on contestants' likelihood of answering a Daily Double question correctly. If players exhibit a hot hand effect, then they should be more likely to get a question correct when "hot" compared to when they are not, even upon controlling for underlying contestant ability and the state of the game. Conversely, if no such effect exists, then the probability of a correct answer should not be able to predicted by whether or not a contestant is "hot."

Let C_{ijt} denote a binary indicator of whether or not contestant i answered a clue

[10] This functional form is commonly employed when diminishing returns are relevant, for example in much of the literature studying the effects of advertising on demand.

[11] Given that each game is approximately 60 clues (61 at maximum if we include the final round), these choices of M capture "recent" performance to a suitable extent without approaching full detailed histories. We also note that the bulk of M-average observations become restricted to smaller ranges for higher M, since it becomes more and more rare for a player to get most or all of the last M clues correct as M increases.

correctly at time period t in game j, Z_{ijt} be a vector of controls, and η_{ijt} an error term. We define each time period to consist of exactly one question for a given game and also write S_{ijt} to represent the hot hand state variable, where its value for a given contestant i at time t in game j corresponds to one of the "hot"-ness measures defined in Section 3.2.

We include the following covariates in the control vector Z_{ijt}: the contestant's current score, an indicator for whether the question is during the Double Jeopardy round, an interaction term between score and round indicator, the contestant's lead amount if leading, the contestant's deficit amount if behind, and contestant fixed effects. We also control for whether or not a Daily Double is in the first or last 10 questions of a game, since such questions may elicit certain unique tendencies due to their proximities to the beginning or end of the game respectively. Finally we control for the difficulty level of the clue, which ranges on a scale of 1–5, with 1 corresponding to the top row on the game board and 5 the bottom row, since question difficulty increases for higher value clues.

To test for a hot hand effect, we then use the specification

$$Pr(C_{ijt} = 1) = \gamma_S S_{ijt} + \gamma Z_{ijt} + \eta_{ijt}. \tag{1}$$

We make a few additional assumptions in order for our model estimates to provide consistent estimates for the desired effects. First, we assume that contestants' baseline abilities remain constant over time. While learning is certainly possible and presents a potential confound, we note that contestants each participate in a number of practice games against their opponents before the formal game tapings begin, allowing them to adjust to the nuances of gameplay at the recording studio. Similarly to trivia skill, we also assume that a contestant's ability to press their buzzer quickly is included as part

Table 2: Estimates for Determinants of Daily Double Correctness

Variable	Streak	log(1+S)	M = 2	M = 3	M = 4
Hot Hand	0.009*	0.036***	0.078***	0.069***	0.066**
	(0.005)	(0.013)	(0.023)	(0.025)	(0.028)
Difficulty	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Score	−0.005	−0.005	−0.005	−0.005	−0.005
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Round2	0.131**	0.131**	0.135**	0.136**	0.137**
	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)
Score:Rd2	−0.014**	−0.014**	−0.014**	−0.014**	−0.014**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Ld:Ld Amt	0.012***	0.012***	0.013***	0.012***	0.012***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Def:Def Amt	0.002	0.002	0.003	0.002	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Cat. Switch	−0.003	−0.004	−0.009	−0.006	−0.004
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
N	11,573	11,573	11,573	11,573	11,573
R^2	0.501	0.501	0.501	0.501	0.501

This table gives the estimated effects of covariates on Daily Double correctness. All monetary values are in thousands of dollars. The columns correspond to the different measures of "hot"-ness described in Section 3.2.

of their larger overall playing ability. This implies that their baseline buzzing ability also does not change over time, save for short-term fluctuations due to being "hot."

Table 2 estimates the equation via a linear probability model, though the estimates for a probit or other nonlinear model do not substantively differ. We estimate that a one-question increase in correct streak length is expected to increase the probability of a correct Daily Double answer by 0.9%. Similarly, a 1-unit increase in $\log(1 + S)$ is expected to increase the probability of a correct answer by 3.6%, where this is approximately equal to the 0.9% increase predicted by the raw streak variable when evaluated at the mean streak length. For the M-average specifications of "hot"-ness, we estimate that having an M-average of 1 is expected to increase correctness probability by 6.5–8% depending on the chosen M. We also note that though statistically insignificant, switching question category appears to slightly decrease the probability of answering correctly, though the effect is small. Altogether, the estimates in this section indicate that being "hot" is by itself a statistically significant predictor of an increase in the probability of answering a Daily Double correctly.

4.2 Belief in a Hot Hand Effect

A separate but related question to whether a hot hand effect exists is whether or not people *believe* that a hot hand exists. In general, the prior literature has found that the answer is yes; however, it still behooves us to check that this is indeed the case in our setting. For one, our setting is neither sports nor gambling per se (the main focus of prior works), so it is useful to see whether people's belief in a hot hand exists outside these contexts. Furthermore, it is necessary to establish whether or not people believe in a hot hand in order to adequately investigate the important follow-up questions of interest; namely, do people's beliefs align with reality, and in what ways?

To check for contestant belief in a hot hand, we check whether being "hot" increases the amount by which contestants are expected to wager before answering a Daily Double clue. Note that we define belief in a hot hand as a contestant's belief regarding their own performance and not that of other players. Prior works investigating hot hand belief (e.g. Camerer (1989) and Green and Zwiebel (2018)) look at people's beliefs regarding *others'* performances, which is a related but distinct question, since outside observers such as sports bettors or managers may have external information available at the times of their decisions. Given this potential for external assistance, the behaviors observed are not necessarily reflective of these decision-makers' true beliefs. In our setting we estimate the model

$$W_{ijt} = \alpha_S S_{ijt} + \alpha X_{ijt} + \varepsilon_{ijt}. \tag{2}$$

Not that the outcome variable has changed from clue correctness to wager amount compared to the existence model in Section 4.1. Again the covariates we include are the hot hand state variable S_{ijt} and a vector of controls X_{ijt}. In this specification, the controls we include are those likely to affect a contestant's wager outside of a potential hot hand effect: a contestant's current score, an indicator for round, whether or not a clue is in the first or last 10 questions of the full game, the contestant's lead or deficit amount, and contestant fixed effects that serve as a control for contestants' baseline wagering tendencies.[12]

We expect that these covariates have a significant effect on contestants' wagering calculus. For example, a contestant's current score represents the maximum amount that they are allowed to wager.[13] Daily Doubles represent opportunities for contestants to catch up with the leader if behind as well as cement their lead when ahead (or also lose their lead), hence lead and deficit amounts are likely to have a causal effect. The round controls address the possibility that the second round may elicit slightly different strategy as contestants' scores are typically higher and the end of the game is closer. Finally, including contestant fixed effects is important given the potential for differing skill levels and wagering tendencies across players. Frequent watchers of *Jeopardy!* might argue that players often go on streaks within a category and hence may wager more on a Daily Double question due to it being in the same category rather than due to a true hot hand effect. To account for this, we include an indicator for whether or not the current question is in the same category as the previous one.

For interpretation, a positive and statistically significant estimate of α_S implies that a contestant being "hot" results in their wagering more dollars than when they are not "hot." Consequently, this implies that contestants believe in a hot hand on average,

[12]We do not include clue difficulty in the regression, since it should not affect the wager amount through any of the aforementioned variables and for identification reasons discussed in Section 5.1.

[13]One exception is that if their score is below $1000 in the first round or $2000 in the second round, they are allowed to wager up to those amounts respectively.

Table 3: Estimates for Determinants of Wager Amount

Variable	Streak	log(1+S)	$M = 2$	$M = 3$	$M = 4$
Hot Hand	0.098***	0.287***	0.398***	0.453***	0.438***
	(0.016)	(0.041)	(0.070)	(0.078)	(0.086)
Score	0.250***	0.249***	0.255***	0.252***	0.253***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Round2	0.877***	0.884***	0.894***	0.887***	0.886***
	(0.091)	(0.090)	(0.091)	(0.091)	(0.091)
Score:Round2	−0.123***	−0.124***	−0.129***	−0.126***	−0.126***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Lead:Lead Amt.	−0.055***	−0.054***	−0.053***	−0.053***	−0.054***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Def:Def. Amt.	0.165***	0.166***	0.165***	0.166***	0.166***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Cat. Switch	0.087*	0.081*	0.071	0.079*	0.091*
	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)
N	11,573	11,573	11,573	11,573	11,573
R^2	0.267	0.269	0.267	0.267	0.266

This table gives the estimated effects of covariates on player wager amounts for Daily Double questions. All monetary values are in thousands of dollars. The columns correspond to the different measures of "hot"-ness described in Section 3.2.

either consciously or subconsciously.[14]

Results for this specification are provided in Table 3. First, we discuss the estimates for the effects of our control variables. We find that score has a positive and statistically significant effect on wager amount; a 1000 dollar increase in a player's score is expected to increase their Daily Double wager by approximately 250 dollars. Additionally, if a player is in the lead, then a 1000 dollar increase in the magnitude of that lead is expected to decrease their Daily Double wager by approximately 55 dollars. If a player is not in the lead, then a 1000 dollar increase in the magnitude of their deficit is expected to increase their Daily Double wager by 165 dollars.

[14] We also note it is possible that certain covariates omitted from our regression model may be correlated with S_{ijt} and impact W_{ijt}, causing omitted variable bias. In our case this is not problematic, since we simply aim to capture the *total* effect of S_{ijt} on W_{ijt} and attribute any such effect to belief in the hot hand, whether it comes directly through S_{ijt} or indirectly through another variable.

For the round coefficient, we find that Double Jeopardy wagers are expected to increase by approximately 880 dollars relative to those in the first Jeopardy stage. Though it is not immediately obvious why this would be the case, one plausible explanation is that scores, leads, and deficits all increase on average as the game continues, so the scale at which contestants perceive the game scores is higher. The round indicator may also be reflective of anchoring effects or potential nonlinearities of the score effect. Finally, a contestant may use a Daily Double as a way to catch up with the leader, who might be farther ahead in the second round.

We also note that the two findings with respect to lead and deficit magnitudes support Kahneman and Tversky (1979)'s prospect theoretic prediction that people are risk averse to losses and risk seeking for gains. Players who are already behind are willing to risk more money than they otherwise would, implying that the marginal utility of a financial gain is greater than the marginal utility of an equally sized financial loss. Similarly, players who are already in the lead are less willing to risk money, suggesting that the marginal disutility of financial loss is greater than the marginal utility of an equally sized gain when leading. However, we do note that a potential confound to learning about risk aversion from these estimates alone is that there is a discontinuity in payoffs in the scenario where first and second place are tied. We merely point out that the behaviors suggested by prospect theory may be exacerbating these effects.

Next we interpret our estimates of α_S, the coefficient on the "hot hand" regressor as well as our main coefficient of interest. We estimate that wagers are expected to increase by approximately 98 dollars for every unit increase in streak length leading into a Daily Double question. For example, if a player has answered the last 4 questions correctly, they are expected to wager 438 dollars more relative to if the streak was of length 0. The marginal effect of an increase in streak length is not constant for the $\log(1+S)$ specification, capturing the idea of a diminishing effect for higher streaks.

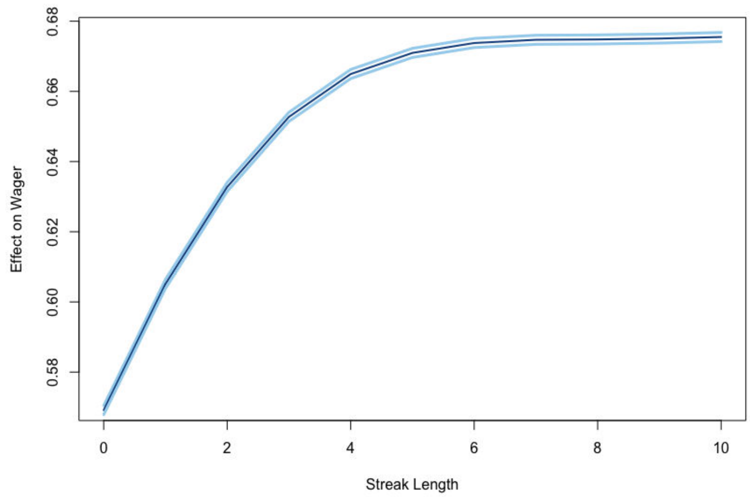

Figure 2: Nonparametric Estimates of Streak Effect on Contestant Wager

This figure visualizes a nonparametric estimate of the effect of streak length on contestant wager with its 95% confidence interval. It is obtained via estimation of the semiparametric partially linear model introduced in Robinson (1988). All other covariates, including contestant fixed effects, enter as in Equation 2. We provide additional detail in Appendix Section A.1.

However, the marginal effect at the mean streak length $\overline{S} = 1.3$ is roughly equal to the standard streak effect suggesting consistency between the specifications, while the marginal effect at other points can be computed analogously.

For the two streak specifications, the assumption of a diminishing effect for higher streak levels appears well-founded. Figure 2 provides a nonparametric estimate of the effect of streak length on daily double wager amount. The strictly increasing but concave pattern suggests that $\log(1 + S)$ indeed provides a more realistic fit than raw streak length.

For the M-average specifications of hotness, we recall that the M-average value falls between 0 and 1, with 1 meaning that a player has answered M out of the last M clues correctly. The interpretation of the $M = 2$ estimate indicates that a player is expected to wager 200 dollars more when they have answered 1 out of the last 2 questions correctly compared to 0 out of the last 2 questions, as well as an additional 200 dollars when answering 2 out of the last 2 questions correctly. The interpretations of the effect of hotness on wager amount for the $M = 3$ and $M = 4$ cases follow analogous reasoning.

Lastly, we note that whether or not a Daily Double question is in the same category as the previous clue does not have a significant effect on the wager amount. This addresses the potential concern that players may choose to sequentially reveal clues in their preferred category – in this case, they may have longer winning streaks as well as higher Daily Double wagers simply because both relate to a preferred category rather than an independent effect of the former on the latter. Our finding that Daily Double questions in the same category as the previous clue do not induce significantly different wagers relative to when they are in different categories helps mitigate this concern.

5 Quantifying a Hot Hand Bias

The results in Section 4.1 suggest that contestants are more likely to answer a question correctly after experiencing recent success, even when suitable controls are added. Furthermore, Section 4.2 indicates that contestants believe in such an effect as reflected through wagers placed on their own performances during gameplay. Consequently, according to our estimates, contestants' beliefs are grounded in reality, drawing direct contrast with the seminal finding in Gilovich, Vallone, and Tversky (1985) and adding to growing evidence that belief in a hot hand may not be fallacious after all.

In this section, we directly address a natural follow-up question motivating much of the literature on the hot hand effect: is belief in the hot hand rational or biased?[15]

[15]More precisely, we ask whether belief in the hot hand is "behaviorally consistent" by seeing whether the increase in wager accurately reflects the true underlying effect. If we assume that the initial strategy without such effects is rational, then the questions are the same.

Namely, do contestants accurately assess the magnitude of the hot hand effect, or are they perhaps systematically over- or underestimating the effect?

Green and Zwiebel (2018) attempted to address this question by analyzing increases in walk rates and other defensive responses when baseball players become hot. However, this approach lends itself to much of the criticism that their paper levied on the prior literature studying the basketball domain, namely that endogenous defensive responses may cause simultaneity bias in estimating magnitudes of or relating to hot hand effects. While it could be argued, like they do, that baseball is less impacted by the issue of endogenous responses, it is still likely that the defensive sides of teams respond to "hot"-ness of batters on the opposing team. Thus, while this context mitigates concerns about a true hot hand effect being masked by the presence of endogenous responses, it does not entirely eliminate it. In particular, these responses may reduce the magnitudes of the measured effects.

Furthermore, even if their empirical strategy holds in its entirety, the baseball setting is not ideal for testing whether humans are able to accurately assess the magnitude of hot hand effects since baseball strategy is largely determined at a team rather than at a player level. Teams hire analysts specifically to develop optimal strategies to address situations such as how a defense should pitch to a "hot" player. Managers then make highly calculated strategic decisions that players must follow. It should then be unsurprising that teams are reasonably proficient in assessing whether or not a player is "hot" and how to approach situations involving that player.

Our setting is particularly well-suited to answer this question. We are able to establish the effect of contestant "hot"-ness on the probability of answering correctly and then see contestants' time-constrained, live responses in the form of monetary wagers of essentially any amount. We can then exploit the fact that players' wager amounts are at least partially impacted by the probability of answering the question correctly and quantify whether contestants' beliefs in a hot hand effect are consistent with its true underlying magnitude.

5.1 Empirical Strategy

In this section, we describe our empirical strategy for identifying and quantifying a hot hand bias in our setting. Ultimately we separate belief in a hot hand into two parts: a behaviorally consistent or "rational" component and a bias term or "irrational" component. A bias term of magnitude 0 implies that people accurately assess the magnitude of the underlying hot hand effect. If it is nonzero, we can quantify the extent to which people over- or underestimate the magnitude of the true effect.

Recall from Equation 1 that we have

$$Pr(C_{ijt} = 1) = \gamma_S S_{ijt} + \gamma Z_{ijt} + \eta_{ijt}$$

where γ_S captures the increase in probability of answering a Daily Double question correctly that can be attributed to being "hot." Furthermore, Equation 2 for contestant wagers is

$$W_{ijt} = \alpha_S S_{ijt} + \alpha X_{ijt} + \varepsilon_{1ijt}$$

such that α_S captures the *total* effect of being "hot" on wager amount. Here we emphasize that the effect is not necessarily a partial effect since we purposefully do not include $Pr(C_{ijt} = 1)$ as a regressor. We posit that the true model for contestant wagers takes the form

$$W_{ijt} = \beta_S S_{ijt} + \beta_C Pr(C_{ijt} = 1) + \beta X_{ijt} + \varepsilon_{2ijt} \tag{3}$$

To understand the mechanics of the estimation when $Pr(C_{ijt} = 1)$ is omitted, we rewrite

Equation 3 as

$$\begin{aligned}
W_{ijt} &= \beta_S S_{ijt} + \beta_C[\gamma_S S_{ijt} + \gamma Z_{ijt} + \eta_{ijt}] + \beta X_{ijt} + \varepsilon_{2ijt} \\
&= [\beta_S + \beta_C\gamma_S]S_{ijt} + \beta X_{ijt} + \beta_C\gamma Z_{ijt} + [\varepsilon_{2ijt} + \beta_C\eta_{ijt}] \\
&= [\beta_S + \beta_C\gamma_S]S_{ijt} + [\beta + \beta_C\gamma_X]X_{ijt} + [\varepsilon_{2ijt} + \beta_C(\eta_{ijt} + \gamma_{M_X Z}M_X Z_{ijt})] \\
&= \alpha_S S_{ijt} + \alpha X_{ijt} + \varepsilon_{1ijt}
\end{aligned}$$

whereby letting P_X, M_X denote the usual projection matrices off of and onto X respectively we can decompose $\gamma Z_{ijt} = \gamma_X P_X Z_{ijt} + \gamma_{M_X Z} M_X Z_{ijt} = \gamma_X X_{ijt} + \gamma_{M_X Z} M_X Z_{ijt}$. We note that $P_X Z_{ijt} = X_{ijt}$ since we defined the control vector Z_{ijt} to be a strict superset of the covariates contained in X_{ijt}.

From the decomposition above, we see that the total effect of being "hot" on wager amount is $\alpha_S = \beta_S + \beta_C\gamma_S$, i.e. it decomposes into the ceteris paribus partial effect as well as an indirect effect entering through $Pr(C_{ijt} = 1)$. We also note that a key assumption of our model is that $M_X Z_{ijt}$ is uncorrelated with ε_{2ijt}. This is sensible in our setting since the only covariate in Z_{ijt} that is not in X_{ijt} is clue difficulty. Logically, clue difficulty should only affect the wager amount indirectly through $Pr(C_{ijt} = 1)$ in that higher clue difficulty implies a lower probability of answering correctly, leading a contestant to wager less all else equal. Thus we assume $M_X Z_{ijt}$ to be uncorrelated with the error term ε_{2ijt}, allowing for consistent estimation of Equation 3 provided that we observe all covariates.

The critical issue to address is that we do not observe $Pr(C_{ijt} = 1)$. Hence, to correct for omitted variable bias and properly identify β_S, we need to either find an instrument for S_{ijt} or proxy for $Pr(C_{ijt} = 1)$. Given our model estimates for hot hand existence in Section 4.1, we can proxy for the unobserved probabilities via

$$Pr(\widehat{C_{ijt} = 1}) = \widehat{\gamma_S} S_{ijt} + \widehat{\gamma} Z_{ijt}$$

We can then estimate Equation 3 using the probability proxies in place of the unobserved true values.

An important characteristic of these proxies is that they are exogenous to ε_{2ijt} by construction since clue difficulty affects W_{ijt} solely through the proxies and has nonzero effect on the proxies. With our setting, exogenous proxies are necessary for the separate identification of (β_S, β_C). We also note that our strategy is not an instrumental variables estimation per see, though we require the proxies to satisfy relevance and exclusion restrictions much like a good instrument.

Intuitively, our strategy exploits omitted variable bias by comparing the difference between the partial effect of being "hot" to the total effect of being "hot" on contestant wagers. We then stipulate that the difference in these amounts enters through the probability of answering correctly, which likely has a causal effect on wager amount and is omitted in Equation 2. The partial effect of being "hot" on a contestant's wager amount is then interpreted as a hot hand "bias." Provided that our model is correctly specified, it is the behaviorally inconsistent effect of being "hot" on a contestant's expected wager: it is an additional increase (or decrease) in wager amount that cannot be explained by the increased probability of answering correctly that comes with an underlying hot hand effect on the probability of answering a question correctly.

5.2 Results

The empirical strategy outlined in Section 5.1 separates the "rational" component of contestant belief in a hot hand effect from a bias term. We note that in this setting we do not define "rational" to have the standard economic interpretation in terms of a strict maximization of some underlying individual utility function. Rather, we take the baseline wager in the absence of any hot hand effects as given, and say that a person is rationally interpreting the hot hand effect if the increase in wager attributed to being "hot" exactly reflects the increase in probability that actually comes with being "hot." In this sense, it may be more intuitive to think of the "rational" hot hand component as being *behaviorally consistent* rather than rational in the conventional sense. The unexplained remaining increasing in wagers attributed to being "hot" is interpreted as the hot hand bias term. We call this the "irrational" component of the hot hand effect.

We have two methods for consistently estimating the rational component. The first is to use the estimate $\widehat{\alpha_S} - \widehat{\beta_S}$ and the second is to use $\widehat{\beta_C}\widehat{\gamma_S}$. Similarly, the bias term can be estimated via either $\widehat{\alpha_S} - \widehat{\beta_C}\widehat{\gamma_S}$ or $\widehat{\beta_S}$. Asymptotically the approaches are equivalent provided that our model assumptions hold. In practice, in our data we find that the estimates differ by less than 5% regardless of which method is used. Nevertheless, since the latter estimates are more direct for both, we report those in Table 4.

We see that in general, our estimates for both components are positive and statistically significant. Note that the interpretation of any positive bias term at all is that contestants' belief in the hot hand is not fully rational in that a residual hot hand effect still causes an increase in expected wager, even accounting for the boost in the probability of correctness that the underlying effect provides. In other words, contestants overestimate the magnitude of the effect. This is possibly due to overadjustment of baseline wagers in response to detection of a hot hand effect, though we do not take a stance on the behavioral or neurological mechanisms that may generate these effects.

Concretely, the bias term is 3–8 times as large as the irrational component across varying specifications of "hot"-ness. This suggests a significant and nonneglible degree of positive bias in contestants' belief in their own hot hand. Since the ratios are greater than 1 across specifications, we deduce that contestants do not simply overestimate the hot hand effect by a smaller or equally sized amount than the true effect would suggest; rather, they substantially overestimate it.

We cannot answer from our setting whether this overestimation would persist if an underlying hot hand effect did not exist in reality. Furthermore, results from the prior literature would appear to suggest a mixed conclusion. In domains where humans are indirectly in control of their own "hot"-ness in that they perform the actions themselves (e.g., sports), a hot hand has been found when proper controls are introduced. If there is a domain with no hot hand effect where humans control the underlying action that they try to guess the success likelihood of, it would be interesting to see whether this bias persists due to overconfidence or an alternate behavioral mechanism, or whether it only comes about once a true underlying effect exists.

While we cannot rule out the possibility that the bias component contains certain residual effects caused by the functional form assumptions that our linear model imposes, we note that the rational component is susceptible to these same confounds. Hence there is no reason to interpret a higher hot hand bias than the rational term as a model artifact. Indeed, when we limit the analysis to highly successful players with for example 10 or more wins, we find the bias term to be weakly negative, indicating an underestimation of the true existence effect (which still persists).

5.3 Hot Hand Effects Across Skill Levels

Our prior analyses analyze the full data with all contestants. A natural follow-up question is whether the findings with respect to hot hand effects change across skill levels. Particularly since our results suggest the existence of a hot hand bias, it is possible that for example, more skilled players will exhibit different characteristics pertaining to the bias.

Table 4: Estimates of Rational and Bias Components of Hot Hand Belief

Hot Hand Measure	HH Bias	Rational	Ratio
Streak Length	0.087	0.011	7.941
log(1 + Streak Length)	0.244	0.040	6.096
Average of Last $M = 2$	0.303	0.089	3.422
Average of Last $M = 3$	0.370	0.081	4.590
Average of Last $M = 4$	0.361	0.081	4.475

This table reports effect sizes for the decomposition of hot hand effect on wager amount into two components. The "Rational" column provides estimates of the rational (or behaviorally consistent) component that enters indirectly through $Pr(C_{ijt} = 1)$ and the "HH Bias" column provides estimates of the residual bias term, which captures the irrational component. The "Ratio" column is the ratio of the bias component to the rational component.

In Figure 3, we test whether the ratio of the irrational (bias) to rational components of the hot hand effect varies across skill level. For both panels in the figure, we vary the minimum number of wins necessary for a contestant to be included in the analysis. We then compute both the ratio (shown in the left panel) and the total hot hand effect (shown in the right panel) as described in Section 5.1. We use the number of wins as a proxy for contestant skill, positing that more skilled contestants win more games before losing.[16] While this proxy is imperfect in that it is certainly possible for a highly skilled player to play one game, lose, and not return for another game, the general trend should certainly see more skilled players enjoy more wins on the show.

We see that both metrics are decreasing in skill. The total hot hand effect on contestant wager remains positive but sees its magnitude cut roughly in half as we remove contestants with less than 10 lifetime wins. Nonetheless, there is still a significant and positive hot hand effect on contestant wager, even for the highest-skilled group.

The left panel suggests that this can be attributed to the magnitude of the bias decreasing for more skilled players. Recalling that the irrational-to-rational component ratio is 0 when contestants reflect a perfect assessment of the true hot hand magnitude in

[16]Once a contestant loses on *Jeopardy!*, they no longer return for any more regular season games.

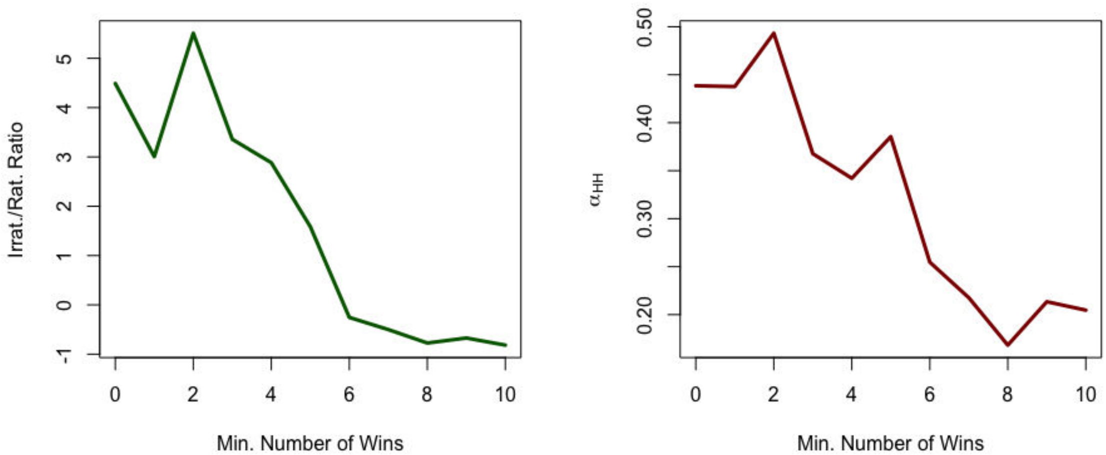

Figure 3: Hot Hand Bias Metrics vs. Skill

This figure analyzes how players of varying skill levels manifest hot hand effects differently. For both panels, the x-axis shifts the minimum number of wins for players included in the calculation. The left panel plots the irrational-to-rational component ratio, while the right panel plots the total hot hand effect estimate $\widehat{\alpha_S}$.

their wagers, we see that the ratio decreases for more skilled contestants and eventually becomes negative for contestants who have won 6 or more games.

This suggests that the most skilled contestants slightly underestimate the magnitude of the underlying hot hand. From this analysis alone we cannot deduce whether there is a causal or correlative relationship between competitive success and hot hand effect magnitudes. One potential explanation for the decreasing bias is that better players are more well-versed in empirically optimal betting strategies while being more familiar with their own playing tendencies due to a larger history of games. Nonetheless, whether higher skill decreases bias or decreased bias leads to greater success remains open for additional exploration.

6 Behavioral Foundations of Hot Hand Effects

6.1 Time Dependence of Hot Hand Effects

In prior sections, we do not take a stance on what might generate an underlying hot hand effect. Given the debate as to whether or not the effect even exists in the first place, the prior literature provides little direction on this fundamental question. Green and Zwiebel (2018) estimate hot hand effects via a state-based Markov model, inducing an oscillation between two states: hot and normal. Gilovich, Vallone, and Tversky (1985) suggest confidence as one possible underlying mechanism; indeed, in the basketball setting, it seems plausible that a player who has recently made shots would be less likely to second-guess another shot, perhaps leading to a more natural shooting stroke or better shot selection. The results of Lantis and Nesson (2021) point to the possibility that hot hand effects are more likely to appear in settings where an action is taken repetitively and consecutively. For the most part, these theories remain speculative, and furthermore it is not clear if they may directly translate to the *Jeopardy!* setting.[17]

Despite the existence of a hot hand effect in our data, we are also unable to determine a mechanism per se from the available variation. However, we provide direction on this question by analyzing whether there is a temporal component to the existence of hot hand effects. To accomplish this, we exploit the regular presence of commercial breaks during gameplay: each *Jeopardy!* show presents clues to contestants, in rapid succession, one after another. Anecdotally, the flow of gameplay is rhythmic in nature, despite variation in questions (within games) and contestants (across games). The only interruptions to this flow are brought about by three 2-minute commercial breaks: one after the 15th clue, one after the first round, and one after the second round. The first two commercial breaks are directly relevant to the Daily Double data we use and

[17] It is also certainly possible that confidence is relevant in explaining a hot hand effect in the basketball setting, while an entirely different mechanism produces one in the *Jeopardy!*.

hence provide natural treatments for interruptions of gameplay. We find that a hot hand effect on performance is only present for clues that directly follow each other; in contrast, for clues that directly or closely follow a commercial break, whether or not a contestant is "hot" has no effect on performance.

First, we test whether there are systematic differences in hot hand or other effects on contestants' ability to answer Daily Double questions correctly by splitting the data into two subsets: one containing questions that are within the five directly following a commercial break and one containing all other questions. The former contains 1103 of the original 11573 data points. For ease of comparing significance levels, we take a random subset of size 1103 from the subset containing all other questions.[18] We then estimate Equation 1 on each of the two subsets and report estimates for the hot hand effects and intercepts in Table 5. Specifications (1) and (2) contain results from the estimation for the first subset (i.e., questions that follow commercial breaks), while specifications (3) and (4) show results for the second (i.e. questions that precede commercial breaks). We also note that (1) and (3) uses streak length to proxy for the hot hand variable S_{ijt}, while (2) and (4) use the M-average measure. In this case we use $M = 3$, though the results do not substantively change for similar M.

We see that in the first two columns that use data from clues that come soon after commercial breaks, the hot hand variable S_{ijt} does not have any significant effect on question correctness, unlike the results with the full data in Table 2. This suggests that hot hand effects disappear in the aftermath of commercial breaks. However, in the next two columns, both measures of S_{ijt} have the indicative significantly positive effect on question correctness; the effect is particularly strong when using M-average as the measure.[19]

[18] Our sampling strategy ensures that we have an equal number observations for the two sets of questions being compared, giving maximum opportunity for failure to reject the null hypothesis of no time dependence.

[19] Recall from earlier that a variety of reasons suggest M-average to be a more direct measure of

To check for robustness, we can run an alternative model, again with Daily Double correctness as the outcome variable. Letting $AB_{ijt} = 1$ if the question answered by player i at time t is within the five immediately following a commercial break and 0 otherwise, we estimate the equation

$$Pr(C_{ijt} = 1) = \gamma_S S_{ijt} + \gamma_{AB} AB_{ijt} + \gamma_{S:AB} S_{ijt} \times AB_{ijt} + \gamma Z_{ijt} + \eta_{ijt} \tag{4}$$

The coefficient $\gamma_{S:AB}$ captures the increase or decrease, if any, in the hot hand effect depending on if questions come soon after a break versus if they do not. The theory that hot hand effects are acutely time-sensitive would be supported if we reject the null hypothesis of $\gamma_{S:AB} = 0$ in favor of the one-sided alternative $\gamma_{S:AB} < 0$.[20] In line with the hypothesis implied by the results of the first four columns, we find that γ_S is statistically significant and positive, while $\gamma_{S:AB}$ is statistically significant and negative. Furthermore, the sum $\widehat{\gamma_{AB}} + \widehat{\gamma_{S:AB}}$ is less than zero, confirming that there is a significant net decrease in the probability of answering a Daily Double correctly in the aftermath of a commercial break compared to the standard case.

A natural follow-up question to the analysis in this section is how (if at all) contestant beliefs account for the dissipation of hot hand effects immediately following commercial breaks. In Table 13 of Appendix Section A.2, we show that contestant beliefs do not reflect acknowledgment of the reduction in $P(C_{ijt} = 1)$ following commercial breaks, indicating that contestants' are unaware of any potential internal shifts that may cause dissipation of the true effect. We further discuss potential cognitive foundations of hot hand belief in Section 6.2.

"hot"-ness than streak length

[20]Note that obtaining $\gamma_{S:AB} > 0$ is also possible and would imply that hot hand effects become stronger after commercial breaks. Such an outcome would likely be irreconcilable with our proposed theory.

Table 5: Impact of Commercial Breaks on Contestant Performance

Variable	(1)	(2)	(3)	(4)	(5)
Streak	−0.003 (0.022)		0.027** (0.013)		
M-Average		−0.025 (0.107)		0.209*** (0.066)	0.087*** (0.027)
After Break					0.046 (0.041)
M-Avg: After Break					−0.113* (0.066)
N	1,812	1,812	1,812	1,812	11,573
R^2	0.840	0.840	0.558	0.561	0.501

This table reports estimates from specifications that illustrate the differences in true hot hand effect magnitudes depending on whether the relevant questions closely follow a commercial break. The three specifications are as discussed in the text; recall that (1) and (2) use data from all questions within 5 of a commercial break; (3) and (4) use a randomly selected sample of all other questions such that it is the same size as that used in (1) and (2); and (5) uses the full data. The first four specifications estimate analogues of Equation 1, while the fifth specification estimates Equation 4. Here, M-average is computed using $M = 3$.

6.2 Analysis of Player Professions

Jeopardy! contestants come from a wide variety of professions, which we observe in our data. Examples include "software engineer," "librarian," and "social science professor." In this section, we provide evidence that players in professions with higher analytical demands, and in particular those requiring greater quantitative training, are highly correlated with lower levels of hot hand bias.

We hired multiple research assistants to designate the most common 500 professions in our data as "quantitative" or not (on a 0-1 scale), as well as "analytical" or not. Since these ratings are inherently subjective, we averaged the provided ratings, yielding an average score between 0-1 for each profession and each measure (quantitative and analytical). Some examples of these scores include:

We then separate the Daily Double clues into two subsets: one with data from contestants in professions with quantitative scores higher than 0.5 and one with the

Table 6: Example Quantitative and Analytical Profession Scores

Profession	Q-Score	A-Score
Bartender	0.00	0.00
Economist	1.00	1.00
Homemaker	0.50	0.50
Lawyer	0.33	1.00
Police Officer	0.33	0.33
Software Engineer	1.00	1.00
Tutor	0.67	0.83
Writer	0.00	0.67

Examples of quantitative ("Q-score") and analytical ("A-Score") averages for select professions.

Table 7: Profession-Stratified Estimates of Hot Hand Bias

Hot Hand Measure	Q-High	Q-Low	A-High	A-Low
Streak Length	0.892	7.115	1.705	2.695
log(1 + Streak Length)	0.850	6.771	1.342	3.876
Avg. of Last $M = 2$	0.850	2.715	0.901	1.881
Avg. of Last $M = 3$	0.817	2.753	1.117	1.076
Avg. of Last $M = 4$	1.032	0.959	0.928	-0.230
N	2,510	3,505	4,481	1,613

This table reports estimates of hot hand bias (the same measure as reported in the "Ratio" column of 4) separated by contestant professions. "Q-High" and "A-High" indicates estimates for contestants from more quantitative and analytical professions respectively; "Q-Low" and "A-Low" are defined analogously.

remaining data. We then estimated the hot hand bias ratio as estimated in Section 5.2. In the table, "Q-High" indicates estimates for contestant professions with quantitative and analytical scores above 0.5, and vice versa for "Q-Low." Separately, we repeat this procedure again using analytical cutoffs rather than quantitative cutoffs, with "A-High" and "A-Low" defined symmetrically to the quantitative case. The results for both procedures are shown in Table 7.

We see that the hot hand bias estimates for contestants from professions with lower quantitative scores are similar to those shown in Table 4 for the full data. However, the bias estimates decrease markedly for contestants from professions with higher quantitative scores to almost 0. The directionality of this contrast is the same for the high versus low analytical profession groups but is less pronounced.

The natural takeaway is that there is a positive correlation between higher degrees of hot hand bias and professions with lower quantitative demands. We cannot further distinguish the nature of this relationship given the variation in our data. It is highly plausible that contestants from more quantitative or analytical professions receive more training in quantitative methodologies, leading such contestants to make to more accurate split-second quantitative judgments (note that players are only given 5–10 seconds to declare a wager, which we use to estimate the magnitude of contestants' biases). It is also plausible that are selection effects into certain professions that also correlate with contestants having little to no hot hand bias. In any case, the correlation here suggests that in domains with both and effects, players' quantitative or general analytical abilities have an inverse relationship with hot hand bias.

Looking again at the estimates in Table 7, we note that the estimates appear to coarsen for higher values of M in the M-average specification of the hot hand state variable S_{ijt}. A likely explanation for this is that as M increases, it is less and less likely to observe data points with high S_{ijt}. Since noise gets introduced into the estimation with the subjective measurements of professions' quantitative and alalytical scores as well as the hot hand measure, it is likely that there is insufficient data to adequate estimate the bias for high values of M. Nonetheless, we continue to report results for $M = 2$ through $M = 4$ as in previous sections.

7 Discussion

We analyzed detailed play-by-play data from the game show *Jeopardy!*, focusing in particular on the interation between contestants' wagering behavior and overall performance on Daily Double questions. To measure whether a contestant's performance is "hot" or not, we followed Green and Zwiebel (2018)'s conceptualization of a hot hand effect as "short-term predictability in performance." We use a variety of measures including streak length (entering both parametrically and nonparametrically) and accuracy in recent clues to test for belief in a hot hand by checking whether being "hot" has an effect on a contestant's expected wager. We found that there is a positive and statistically significant effect: contestants; for instance, a contestant's wager is expected to increase by 98 dollars for each additional question in a row that they answer correctly leading into the Daily Double question. This confirms findings from the prior literature that people typically exhibit belief in a hot hand. It is worth noting that we specifically measured a person's belief in a hot hand effect on their *own* performance; many prior studies such as Camerer (1989) analyze people's beliefs in a hot hand effect on *others'* performances. The distinction is important particularly if biased belief in a hot hand is a result of overconfidence and remains open for future work.

We then tested whether this belief has any basis, the topic of much debate since Gilovich, Vallone, and Tversky (1985)'s seminal paper claimed that a hot hand effect did not exist despite people potentially believing in one. In our setting, we found that being "hot" does directly lead to a positive and significant effect on question correctness. Given that contestants believe in a hot hand and that such an effect indeed exists, we then test whether their beliefs are rational. We find that they are not, and that contestants systematically overestimate the true hot hand effect by a multiplicative factor of 3–8. Moreover, this bias dissipates for increasingly successful contestants as measured by number of wins on the show.

Finally, we analyzed potential causal mechanisms of a hot hand effect on performance. We found that the effect disappears after commercial breaks and reappears shortly thereafter, suggesting that there is a time-dependent component to the existence of a hot hand effect. In other words, a short break before a question predicts that a hot hand effect will not exist, regardless of whether or not the contestant has been on a "hot" streak of answering questions correctly. Furthermore, we find that biased belief in the hot hand is more pronounced for contestants from profession with less quantitative or analytical demands, suggesting that incorrect personal assessments of a hot hand may come down to ability to make split-second quantitative judgments accurately rather than detecing something more innate or endogenous.

Our study's primary contributions are as follows: first, we provide a simultaneous test of both existence of and belief in a hot hand effect in a single setting. Secondly, we develop a framework that allows us to distinguish between rational and irrational components of wager increases that arise due to hot hand effects. Explicitly, we find that contestants tend to over-wager the behaviorally consistent amount by a factor of 3–8 times and that this bias is significantly smaller in more successful contestants. Finally, we shed light on potential mechanisms that might generate a hot hand as well as why contestant beliefs may be biased relative to the true underlying hot hand effect; we analyze both performance after commercial breaks as well as discrepancies in bias estimates stratified by contestants' professions.

Previously, the hot hand phenomenon has been thought of as an illusion or "fallacy." Recent studies have begun to indicate that this is not the case; specifically, hot hand effects on performance may exist in activities with frequent and repetitive components. Our work strongly supports this notion and goes further to show that despite the existence of a hot hand effect on performance, players significantly overestimate its magnitude. This suggests that the hot hand phenomenon is much more accurately conceptualized as a behavioral bias, where the bias may dissipate in a variety of circumstances such as increased player experience, quantitative training, or automated decision making. Our findings also have significant implications for managerial decisions pertaining to stock investments and analyst hirings, optimal decisions in sports, and more generally any setting where repetitive human action drives the underlying stochastic process. Going forward, an important goal for research on this phenomenon is to provide a comprehensive, detailed account of a generating mechanism for hot hand effects. We investigated a number of possibilities in this paper and continue to explore them in ongoing work.

References

Bocskocsky, Andrew, John Ezekowitz, and Carolyn Stein (2014). "Heat Check: New Evidence on the Hot Hand in Basketball". Working Paper.

Camerer, Colin F. (1989). "Does the Basketball Market Believe in the 'Hot Hand'?" In: *American Economic Review* 79(5), pp. 1257–1261.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue (2016). "Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires". In: *Quarterly Journal of Economics* 131(3), pp. 1181–1242.

Croson, Rachel and James Sundali (2005). "The Gambler's Fallacy and the Hot Hand: Empirical Data from Casinos". In: *Journal of Risk and Uncertainty* 30, pp. 155–209.

Dorsey-Palmateer, Reid and Gary Smith (2004). "Bowlers' Hot Hands". In: *The American Statistician* 58(1), pp. 38–45.

Gilovich, Thomas, Robert Vallone, and Amos Tversky (1985). "The Hot Hand in Basketball: On the Misperception of Random Sequences". In: *Cognitive Psychology* 17, pp. 295–314.

Green, Brett and Jeffrey Zwiebel (2018). "The Hot-Hand Fallacy: Cognitive Mistakes or Equilibrium Adjustments? Evidence from Major League Baseball". In: *Management Science* 64(11), pp. 5315–5348.

Guryan, Jonathan and Melissa S Kearney (2008). "Gambling at lucky stores: Empirical evidence from state lottery sales". In: *American Economic Review* 98(1), pp. 458–473.

Jin, Lawrence (2018). "Don't Overshoot: Evidence of Hot Hand Bias from World Darts Championship". PhD Dissertation.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision Under Risk". In: *Econometrica* 47(2), pp. 263–291.

Lantis, Robert M. and Erik T. Nesson (2021). "The Hot Hand in the NBA 3-Point Contest: The Importance of Location, Location, Location". NBER Working Paper.

Metrick, Andrew (1995). "A Natural Experiment in Jeopardy!" In: *American Economic Review* 85(1), pp. 240–253.

Miller, Joshua and Adam Sanjurjo (2018). "Surprised by the Hot Hand Fallacy? A Truth in the Law of Small Numbers". In: *Econometrica* 86(6), pp. 2019–2047.

Narayanan, Sridhar and Puneet Manchanda (2012). "An empirical analysis of individiual level casino gambling behavior". In: *Quantitative Marketing and Economics* 10(1), pp. 27–62.

Rabin, Matthew (2002). "Inference by Believers in the Law of Small Numbers". In: *Quarterly Journal of Economics* 117(3), pp. 775–816.

Rabin, Matthew and Dimitri Vayanos (2010). "The Gambler's and Hot Hand Fallacies: Theory and Applications". In: *Review of Economic Studies* 77(2), pp. 730–778.

Ritzwoller, David M. and Joseph P. Romano (2022). "Uncertainty in the hot hand fallacy: Detecting streaky alternatives to random bernoulli sequences". In: *Review of Economic Studies* 89(2), pp. 976–1007.

Robinson, Peter M. (1988). "Root-N-Consistent Semiparametric Regression". In: *Econometrica* 56(4), pp. 931–954.

Tversky, Amos and Daniel Kahneman (1971). "Belief in the law of small numbers". In: *Psychological Bulletin* 76(2), pp. 105–110.

A Appendix

A.1 Nonparametric Estimation of Streak Effect

In this section we provide additional detail on Robinson (1988)'s semiparametric model that allows us to estimate the functional form of the streak effect on wager amount in Figure 2. Relaxing Equation 2 to allow for additional flexibility in the streak effect, we write

$$W = m(S) + X\alpha + \varepsilon,$$

where $m(\cdot)$ is a possibly nonlinear function of streak length; for simplicity we notate the generic hot hand variable S_{ijt} here as S_{ijt} for greater specificity. Then we have

$$E[W|S] = m(S) + E[X|S]\alpha$$
$$W - E[W|S] = (X - E[X|S])\alpha + \varepsilon$$

Letting $\tilde{W} = W - E[W|S], \tilde{X} = X - E[X|S]$, we have that the OLS estimator of α is given by $\hat{\alpha} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}$ where $E[W|S], E[X|S]$ have been replaced by their respective nonparametric estimates $\hat{E}[W|S], \hat{E}[X|S]$. In our case, we estimate the conditional means using the Nadaraya-Watson estimator

$$\hat{E}[Z|S = s] = \frac{\sum_{j=1}^{N} K(s - s_j)z_j}{\sum_{j=1}^{N} K(s - s_j)},$$

where $K(\cdot)$ is a Gaussian kernel. Finally, upon estimating $\hat{\alpha}$, $m(S)$ can be recovered via the relation $\hat{m}(S) = \hat{E}[W|S] - \hat{E}[X|S]\hat{\alpha}$. To obtain 95% confidence bounds, we bootstrap this procedure over 200 iterations.

A.2 Post-Break Shifts in Hot Hand Belief

Table 8: Impact of Commercial Breaks on Hot Hand Belief

Variable	(1)	(2)	(3)	(4)	(5)
Streak	0.140**		0.035		
	(0.063)		(0.039)		
M-Average		0.133		0.287	0.400***
		(0.313)		(0.201)	(0.084)
After Break					−0.019
					(0.125)
M-Avg: After Break					0.313
					(0.203)
N	1,812	1,812	1,812	1,812	11,812
R^2	0.877	0.875	0.661	0.662	0.656

This table reports estimates from specifications that illustrate the differences in true hot hand effect magnitudes depending on whether the relevant questions closely follow a commercial break. The three specifications are as discussed in the text; recall that (1) and (2) use data from all questions within 5 of a commercial break; (3) and (4) use a randomly selected sample of all other questions such that it is the same size as that used in (1) and (2); and (5) uses the full data. The first four specifications estimate analogues of Equation 1, while the fifth specification estimates Equation 4. Here, M-average is computed using $M = 3$.

A.3 Robustness Check: Analysis for Players with 1+ Wins

Table 9: Estimates for Determinants of Wager Amount (Players with 1+ Wins)

Variable	Streak	log(1+S)	$M = 2$	$M = 3$	$M = 4$
Hot Hand	0.095***	0.286***	0.399***	0.440***	0.438***
	(0.018)	(0.046)	(0.079)	(0.088)	(0.096)
Score	0.256***	0.256***	0.262***	0.259***	0.259***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Round2	1.047***	1.052***	1.066***	1.062***	1.062***
	(0.108)	(0.108)	(0.108)	(0.108)	(0.108)
Score:Round2	−0.140***	−0.141***	−0.146***	−0.144***	−0.143***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Lead:Lead Amt.	−0.047***	−0.045***	−0.044***	−0.045***	−0.046***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Def:Def. Amt.	0.194***	0.194***	0.194***	0.194***	0.194***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Cat. Switch	0.086	0.081	0.071	0.079	0.090
	(0.053)	(0.053)	(0.054)	(0.053)	(0.053)
N	7,244	7,244	7,244	7,244	7,244
R^2	0.261	0.262	0.261	0.261	0.260

This table gives the estimated effects of covariates on player wager amounts for Daily Double questions. All monetary values are in thousands of dollars. The columns correspond to the different measures of “hot”-ness described in Section 3.2.

Table 10: Estimates for Determinants of Daily Double Correctness (Players with 1+ Wins)

Variable	Streak	log(1+S)	M = 2	M = 3	M = 4
Hot Hand	0.011*	0.043**	0.088***	0.069*	0.065*
	(0.005)	(0.014)	(0.024)	(0.027)	(0.030)
Difficulty	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Score	0.004	0.004	0.004	0.004	0.004
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Round2	0.096**	0.096**	0.097**	0.097**	0.097**
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
Score:Rd2	−0.014*	−0.014*	−0.014*	−0.014*	−0.014*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Ld:Ld Amt	0.005	0.005	0.005	0.005	0.005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Def:Def Amt	0.002	0.002	0.001	0.002	0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Cat. Switch	−0.009	−0.010	−0.015	−0.011	−0.009
	(0.016)	(0.016)	(0.017)	(0.017)	(0.016)
N	7,244	7,244	7,244	7,244	7,244
R^2	0.298	0.299	0.299	0.298	0.298

This table gives the estimated effects of covariates on Daily Double correctness. All monetary values are in thousands of dollars. The columns correspond to the different measures of "hot"-ness described in Section 3.2.

Table 11: Estimates of Rational and Bias Components of Hot Hand Belief (Players with 1+ Wins)

Hot Hand Measure	HH Bias	Rational	Ratio
Streak Length	0.075	0.019	3.982
log(1 + Streak Length)	0.213	0.068	3.127
Average of Last M = 2	0.246	0.142	1.732
Average of Last M = 3	0.322	0.114	2.829
Average of Last M = 4	0.332	0.110	3.006

This table reports effect sizes for the decomposition of hot hand effect on wager amount into two components. The "Rational" column provides estimates of the rational (or behaviorally consistent) component that enters indirectly through $Pr(C_{ijt} = 1)$ and the "HH Bias" column provides estimates of the residual bias term, which captures the irrational component. The "Ratio" column is the ratio of the bias component to the rational component.

Table 12: Impact of Commercial Breaks on Contestant Performance (Players with 1+ Wins)

Variable	(1)	(2)	(3)	(4)	(5)
Streak	-0.005		0.026*		
	(0.021)		(0.013)		
M-Average		-0.057		0.202***	0.092***
		(0.107)		(0.071)	(0.029)
After Break					0.100**
					(0.045)
M-Avg: After Break					−0.158**
					(0.071)
N	1,103	1,103	1,103	1,103	7,244
R^2	0.146	0.147	0.076	0.081	0.043

This table reports estimates from specifications that illustrate the differences in true hot hand effect magnitudes depending on whether the relevant questions closely follow a commercial break. The three specifications are as discussed in the text; recall that (1) and (2) use data from all questions within 5 of a commercial break; (3) and (4) use a randomly selected sample of all other questions such that it is the same size as that used in (1) and (2); and (5) uses the full data. The first four specifications estimate analogues of Equation 1, while the fifth specification estimates Equation 4. Here, M-average is computed using $M = 3$.

Table 13: Impact of Commercial Breaks on Hot Hand Belief (Players with 1+ Wins)

Variable	(1)	(2)	(3)	(4)	(5)
Streak	0.141**		0.026		
	(0.065)		(0.043)		
M-Average		0.075		0.299	0.390***
		(0.329)		(0.229)	(0.094)
After Break					−0.021
					(0.145)
M-Avg: After Break					0.295
					(0.231)
N	1,103	1,103	1,103	1,103	7,244
R^2	0.477	0.469	0.362	0.363	0.427

This table reports estimates from specifications that illustrate the differences in true hot hand effect magnitudes depending on whether the relevant questions closely follow a commercial break. The three specifications are as discussed in the text; recall that (1) and (2) use data from all questions within 5 of a commercial break; (3) and (4) use a randomly selected sample of all other questions such that it is the same size as that used in (1) and (2); and (5) uses the full data. The first four specifications estimate analogues of Equation 1, while the fifth specification estimates Equation 4. Here, M-average is computed using $M = 3$.