Title
Essays in belief formation and decision making

Author
Rao, Justin M.

Publication Date
2010

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays in Belief Formation and Decision Making

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Justin M. Rao

Committee in charge:

Professor James Andreoni, Chair
Professor Nageeb Ali
Professor Uri Gneezy
Professor Craig McKenzie
Professor Joel Sobel

2010

The dissertation of Justin M. Rao is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2010

EPIGRAPH

Grass grows, birds fly, waves pound the sand. I beat people up.

— Muhammad Ali

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like my committee members Nageeb Ali, Uri Gneezy, Craig McKenzie and Joel Sobel, Professors Vince Crawford, Navin Kartik, David Miller and Joel Watson, the Russell Sage Foundation and the UCSD Department of Economics for financial support, my colleagues Benjamin Fissel, Laura Gee, Charles Sprenger and William Peterman and most of all my advisor James Andreoni and parents M. Vikram and Susan Henning Rao.

I would also like to formally acknowledge and thank my co-author for Chapter 1, colleague and dear friend, David Eil.

VITA

2005	B.A. in Economics *summe cum laude*, University of North Carolina at Chapel Hill
2005	B.A. in Philosophy *summe cum laude*, University of North Carolina at Chapel Hill
2010	Ph.D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Essays in Belief Formation and Decision Making

by

Justin M. Rao

Doctor of Philosophy in Economics

University of California, San Diego, 2010

Professor James Andreoni, Chair

This dissertation consists of four separate but related papers. The overarching theme is how decision makers process information, form beliefs and make decisions.

Chapter 1 examines how individuals' beliefs respond to objective information about their ranking on a neutral quality – a meaningless number on a card – or on a quality that has a significant self-image component – intelligence or beauty. For favorable news in the image tasks, subjects respected signal strength and update as "optimistic Bayesians," but they heavily discounted and largely ignored signal strength in processing unfavorable news, leading to noisy posterior beliefs nearly uncorrelated with Bayesian inference. None of these patterns were observed in the control.

Chapter 2 uses standard theoretical concepts to examine the optimality of the shooting decisions of National Basketball Association (NBA) players using a rich data set of shots, collected by watching 60 games and recording the shot conditions and outcomes. Shot timing is modeled as an optimal stopping problem and the prediction of a monotonically declining "reservation expected shot value" with time remaining on the shot clock is confirmed in the data. For shot allocation, the mixed strategy Nash equilibrium also finds strong support in the data.

Chapter 3 studies shot selection in professional basketball to see if players succumbed to the "hot hand fallacy." I find that a majority of the players in the sample significantly changed their behavior in response to hit-streaks by taking more difficult shots, while no player responded to miss-streaks. However, controlling for difficulty, shooting ability is not related to past out-

comes. A quarter of the sample suffered a significant loss in shooting efficiency as a result of their mistaken responses. Consistent performance is linked to heightened individual economic incentives.

Chapter 4 uses a subjective probability sequence (professional basketball shooting) to see if the gambler's fallacy (GF) and the hot hand fallacy (HHF) can both exist within a single agent, as predicted by theoretical models. Across subjects, there is statistically significant GF after short streaks and HHF after long streaks. On the individual level, a transition from GF to HHF is most common, consistent with the theoretical prediction.

Chapter 1

The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself

Abstract

This paper examines how individuals' beliefs respond to objective information about their ranking on a neutral quality – a meaningless number on a card – or on a quality that has a significant self-image component – intelligence or beauty. For favorable news in the image tasks, subjects respected signal strength and update as "optimistic Bayesians," but they heavily discounted and largely ignored signal strength in processing unfavorable news, leading to noisy posterior beliefs nearly uncorrelated with Bayesian inference. None of these patterns were observed in the control. Subjects did not display confirmatory bias, but the results can explain its root cause – disconfirming signals are treated as "bad news" due to the self-esteem loss of realizing one was initially wrong.

1.1 Introduction

Many studies have shown that people have a "psychological defense mechanism" to help maintain positive self-image and confidence. Classic examples are selective memory, biased attributional judgment (good outcomes attributed to skill, bad outcomes attributed to luck) and cognitive dissonance. These strategies accentuate the positive even while, on some level, recognizing the negative. Three related economic concepts, confirmatory bias, self-serving bias and overconfidence, are all rooted in filtering out unfavorable information. A common thread is that all these behaviors hinge on new information being processed differently depending on whether it is favorable or unfavorable to existing beliefs.

However, the literature on how new information is incorporated into existing beliefs (signal processing) has generally shied away from using signals that are inherently good news or bad news. While prominent findings show that people systematically depart from the predictions Bayesian rationality – e.g. the "law of small numbers" [Tversky and Kahneman, 1974] and base rate neglect [Grether, 1980] – the results are difficult to apply to an environment in which beliefs are directly related to self-image and esteem. This is because the experimental protocols typically use an underlying quantity far-removed from self-image, such as the color of balls in an urn.

This paper aims to bridge the gap between these two literatures. We study how people incorporate objective pieces of new information into their existing beliefs when the underlying quantity has direct psychological importance. In this environment signals have intrinsic valence – news is good or bad. The qualities used to generate signal valence were intelligence (*IQ*), as measured by score on an IQ test, and physical attractiveness (*Beauty*), as rated by subjects of the opposite sex during speed dating. As a control, subjects also updated their beliefs under the same information structure on the number (1-10) on a card given to them in a sealed envelope. All subjects participated in one image task and the control task in an experimentally balanced order. For each task, subjects first revealed the distribution of their prior beliefs through a computer interface. Then, they received three rounds of up/down signals. Each signal was an anonymous pair-wise comparison of their rank (out of 10 subjects in their session) to a randomly selected anonymous subject. For *Beauty* there were 10 subjects of each gender and all comparisons were made within gender.

Our main finding is that subjects incorporated favorable news into their existing beliefs in a fundamentally different manner than unfavorable news. For favorable news in the image tasks, subjects tended to respect signal strength and to adhere quite closely to the Bayesian benchmark (albeit with an optimistic bias). In contrast, subjects discounted or ignored signal strength in processing unfavorable news, which lead to noisy posterior beliefs that were nearly uncorrelated with Bayesian inference. These patterns were not observed in the control. For males the bias was strongest for *IQ*, while for females the bias was strongest for *Beauty*. We call this finding the good news-bad news effect.

The differential processing we observe indicates that bad news has an inherent "sting" that selective filtering helps ameliorate. For perfectly informative signals, it is not possible to apply the filtering ex-post, but it may be possible to avoid the signal ex-ante. As perfectly informative signals become more likely to be unfavorable, we expect higher degrees of avoidance. To examine signal avoidance we elicited subjects' willingness-to-pay (WTP) to learn their true rank through a price list equivalent to the Becker-DeGroot-Marchak mechanism. Consistent with our hypothesis, subjects who had received good news in the image tasks were willing to sacrifice some of their earnings to learn their true rank whereas subjects who had received bad news required a subsidy. Since the probability of learning one's true rank increased linearly with WTP, requiring a large subsidy made it highly unlikely or impossible that one's true rank would actually be revealed. In the control WTP did not vary across signal valence.

Our results help explain the underlying cause of confirmatory bias (CB). CB occurs when signals that agree with one's prior belief are incorporated into posterior beliefs in a different manner than signals that disagree with one's prior. Typically this involves underweighting disconfirming evidence and sometimes overweighting confirming evidence.[1] After viewing an equal number of confirming and disconfirming signals (neutral evidence) a CB-agent will view it as supporting the prior belief.

Evidence for CB comes mainly from the psychology literature. Lord et al. (1979) found that subjects, pre-screened for having strong views on the death penalty, who were given the same articles on the preventative value of capital punishment strengthened their conviction regardless their original stance. That is, both sides of the debate became more convinced they were right by the same evidence. Plouss (1991) used partisans in the nuclear safety debate and replicated Lord et al.'s finding. Mahoney (1977) sent a paper to journal referees with the results section altered to either agree or disagree with the referee's published results. He found that referees for which the result was disconfirming were far more critical of the methods section of the paper than referees for which the result was confirming, despite the fact that the method section was identical for both groups. Nickerson (1988) reviews the extensive evidence for CB and Rabin and Shrag (1999) highlights its economic importance .

Subjects in our experiment did not display CB. For instance, a subject who reported believing she had below average beauty tended to overreact to a signal that informed her she was more attractive than a randomly selected comparison subject ("up") and underreact to the opposite signal ("down"). That is, "down," the signal which confirmed her prior, was given less weight than dictated by Bayesian inference. This is precisely the inverse of CB. Initially this result might appear surprising given past findings of CB. However, a careful survey of the literature reveals that CB presents itself when the underlying quantity is not directly tied to self-image. For example, whether or not the death penalty prevents violent crime does not make one feel good or bad intrinsically (unless, of course, one is on death row). Rather it has psychological importance because it is damaging to the ego to admit one was wrong on an issue they care about. Supporting this intuition, CB experiments generally use partisans' beliefs on hot-button debates. A feature of this protocol is that confirming signals are always good news – confirmation and valence are perfectly co-linear. In our design, confirming signals can be good or bad news. We find that what matters is valence, not agreement or disagreement with prior beliefs.

We do not dispute that CB is an empirical regularity. However, our results indicate

[1] Rabin and Shrag (1999) presents a theoretical model of CB that uses this definition as a starting point.

that agreement with prior beliefs is not the underlying cause. We should only expect to observe CB when disconfirming signals are indeed interpreted as bad news — when being right is what matters most, CB is exhibited. This is exactly the structure of past experiments finding CB. An intuitive feature of our signal valence based explanation of CB is that CB is brought into the fold of the psychological defense mechanism, rather than just being a mechanical updating error.

The good news-bad news effect provides an explanation for self-serving bias. Self-serving bias occurs when self-interest shapes what one believes is fair. For instance, in laboratory bargaining games both sides tend to believe strongly their side is in the right even when they are given identical information (see Babcock and Loewenstein (1997) for a review). Information endorsing a social or moral norm that leads to material benefit has positive valence. Accordingly, it receives more weight in the updating process and self-serving bias naturally develops.

Another application of our results is to the empirical finding of widespread overconfidence in self-image related qualities (see Moore and Healy (2008) for a review).[2] Since strong favorable signals induce a large change in beliefs but strong unfavorable signals are heavily discounted, resultant beliefs are biased towards positive self-image and overconfidence develops. Furthermore, the pattern of information acquisition we observe can further fuel overconfidence as people likely to receive good news are the ones seeking out more precise signals.

The question, however, remains: why would a bias such as the good news-bad news effect become part of our psychological constitution? Why is belief accuracy sacrificed in favor of irrational optimism? Many authors have grappled with this question when discussing the psychological defense mechanism more generally. They have argued that the economic and evolutionary value of self-confidence and self-esteem can potentially outweigh the associated costs.[3] Benabou and Tirole (2002) derive conditions on when so-called biases will emerge in a fully rational model. They show, for instance, that overconfidence can help agents commit to initially costly long-term projects that otherwise might not be undertaken. Compte and Postlewaite (2004) show that if self-confidence enhances performance, then rational agents will choose updating rules that are optimistically biased away from Bayesian inference. Brunnermeier and Parker (2005) modify preferences to include "anticipatory utility" and establish a similar result. These papers provide a solid foundation for the existence of valence-dependent updating in human psychology and our results support these models. The models can also be employed to help explain CB. Earlier we argued that CB can be explained by signal valence — these models show that under this explanation CB can result from a fully rational model, while previous models of CB, such as Rabin and Shrag (1998) had to assume it as a "behavioral bias."

The remainder of the paper proceeds as follows. Section 2 details the experimental design. Section 3 provides the inference results. Section 4 provides the information search results. A

[2]See also Alicke and Govorun (2005).

[3]Inference mistakes can indeed have economic consequences. Examples include excessive trading in stock markets [Barber and Odean, 2001] and overconfidence in physicians' diagnostic decisions [Christensen-Szalanski and Bushyhead, 1988].

discussion follows in Section 5. Section 6 concludes.

1.2 Experimental Design

The experiment was conducted at the University of California San Diego Economics Laboratory. There were 7 *IQ* and 4 *Beauty* sessions. *IQ* sessions had 10 subjects each. *Beauty* sessions had 10 male subjects and 10 females subjects. Potential participants were solicited through an online subject database and were told only that the experiment would last 1.5 hours and earnings would be $25 on average. Given the sensitive nature of the experiment (receiving information on intelligence and physical attractiveness) we over-subscribed the sessions to account for people electing not to participate after reading the IRB consent form.[4] Across all sessions, only 2 subjects elected to leave after learning the nature of the experiment. As such, it was necessary to randomly select subjects to leave to pare the sessions down the required number of subjects.

The experiment proceeded in three stages. The first stage of both session types collected the necessary information to rank the subjects on the intelligence or attractiveness. At this juncture subjects were not told why this information was being collected. In *IQ*, subjects took a 25 question IQ test. The questions were taken from a standard Wechsler Adult Intelligence Scale test and involved logic, spatial and verbal reasoning, and general knowledge. Subjects were told the "aims of the experiment depend on you making an honest effort on this IQ test." To incentivize them further 1 question was chosen at random (ex-post) and if they answered this "payment question" correctly $5 was added to earnings.

In *Beauty*, subjects engaged in a speed dating exercise.[5] Each person met 5 subjects of the opposite sex and engaged them in a 4 minute conversation. These meetings were face-to-face with a partition separating each pair. The only restriction placed on the conversations was that they could not reveal identifying information such as their full name or place of residence. To help break the ice, a conversation topics were suggested. During the meetings soft music was played to mimic a real speed dating environment. Most subjects engaged in lively conversation and seemed to enjoy the exercise.

After each meeting, subjects filled out a "speed dating questionnaire" rating their conversation partner (scale 1-10) on three dimensions: friendliness, attractiveness and ambition. These forms were kept in a manila envelope during the meetings and were filled out at separate partitioned work stations to maintain anonymity and encourage honest assessment. Ambition and friendliness were included to reduce the anxiety of filling out the questionnaire. The average

[4]Upon entering the lab subjects were told to carefully read the consent form which told them that they would receive feedback on either their intelligence or physical attractiveness. At that point they were given an opportunity to discreetly exit if they no longer wanted to participate.

[5]The design of the speed dating exercise is similar to Fisman et al. (2006) which used speed dating ratings to examine gender differences in mate selection.

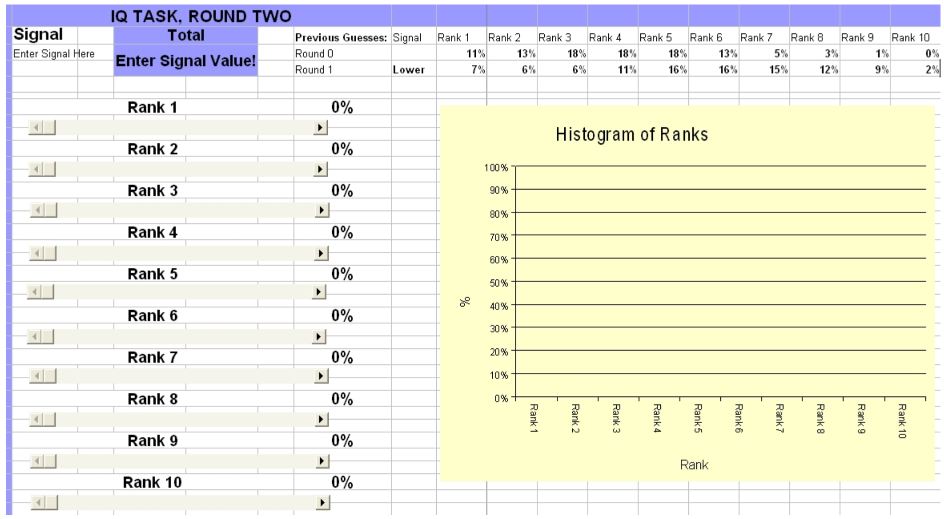


Figure 1.1: Instrument screen-shot.

of the 5 physical attractiveness ratings were used to give each subject a beauty score. After the speed dating exercise, these sessions proceeded as two separate 10 person sessions of each gender.

While stage 2 instructions were read, the scores for the image task in stage 1 were tabulated. The scores ranked the subjects 1 (highest score) through 10 (lowest). Rankings were within gender group for *Beauty*) and ties were broken by a flip of a coin. Subjects were informed of the ranking procedure and understood that the scores gave a strict ordering.

At the beginning of stage 2, subjects were passed a sealed envelope containing a card with a unique number 1-10 written on it. In *Beauty* sessions the numbers were unique within gender group. The number determined rank on the "card task" (*Control*) and again subjects were explained the ranks gave a strict ordering. They were also truthfully told that card rank was determined randomly.

Stage 2 consisted of 2 sets (image and control) of 4 rounds each, enumerated 0 through 3. In round 0, subjects entered their prior belief in percentage terms that they occupied each of the 10 ranks on the task. This was done through a computer interface as seen below in Figure 1.

The percentages always started at 0 for each rank and a graphical representation was provided through a auto-adjusting histogram. The total was given and had to add to 100% to be eligible for payment. Adjustments were made by clicking the arrows or moving the slider-boxes.

Following rounds 0, 1 and 2 each subject's rank was compared to a randomly selected comparison participant (again these were done within gender group for *Beauty* sessions). This comparison participant was anonymous and the comparisons were bi-lateral and unique each round (drawn without replacement). The result of the comparison was conveyed via a message that read either "You are ranked higher" or "You are ranked lower." The message cards were labeled by round and subject ID letter. As the messages were handed out, subjects were reminded each time that "higher" meant closer to the top rank of 1 and "lower" meant closer to the bottom rank of 10. After receiving the message, subjects entered it into their spreadsheet to ensure the message was accurately received. They then entered their new guesses over their rank on the task. The entire history of rank guesses and signals were provided at the top of the screen. The order of the image and control sets was randomized.

Payment for rank guesses was done through the incentive compatible quadratic scoring rule [Selten, 1998]. Following Moore and Heally (2008) subjects were shown the scoring rule formula and told, "Although this formula looks quite complicated, what it means to you is simple. You make the most money on average by honestly reporting your beliefs of the probability you occupy each rank. The formula rewards accuracy in a way that the way to maximize your average winnings is to report honestly." Eliciting beliefs in a laboratory setting indeed a challenge and many authors have highlighted the shortcomings of the quadratic scoring rule. For instance, risk aversion can induce median bias [Holt, 1986]. However, our protocol is designed to identify the processing differences between good and bad news. That is, we will be using comparisons that condition the elicitation mechanism used.

Following completion of stage 2, subjects were given an opportunity to learn their true rank on both the control and image tasks. This was done separately for each task through a price list equivalent to a discretized version of the Becker-DeGroot-Marschak (BDM) mechanism [Becker et al., 1964]. This allowed us to capture subjects' willingness to pay for complete information. Each item of the price list had a choice of the following form:

> Do you prefer (A) Receiving \$$x$ and learning your true rank on the [IQ/Beauty or Card] task or (B) Receiving \$0 and not learning your true rank?

x ranged from -7.00 to 7.00 in \$1.00 increments. When $x < 0$ the language was changed to "paying $-x$" (e.g. if $x = -3$ it read "paying \$3.00"). There were 15 choices for each task. A random number was drawn for each task to determine which decision would be executed. Subjects were told they should answer each question as if it were going to be executed.[6] If the executed choice led to the subject learning their true rank on a given task, then they were given an "acknowledgment form" on which they initialed by their rank indicating they indeed learned the truth. In *Control* they also got to open the envelope. To ensure choice anonymity, if the executed choice did not lead to a subject learning their true rank then they received a blank acknowledgement form.

Each session lasted about 1.5 hours and subject earnings averaged \$23.33. Full instructions can be found in the appendix.

[6] Note that this procedure is equivalent to the BDM but does not require the lengthy and confusing explanation of why subjects have an incentive to report their true willingness to pay.

1.3 Inference Results

The purpose of the design was to have objective signals identical across the image and control treatments. This allows for an easy comparison of the patterns of observed in the neutral *Control* to *Beauty* and *IQ* for which subject rank is tied to self-image. Also, conditional on priors collected in round 0, the design allows us to calculate closed form Bayesian posteriors for rounds 1-3, which are used as the normative benchmark. Objective signals also have the attractive property that the results of the experiment cannot be explained by subjects simply "inverting" (or misinterpreting) some unfavorable signals.

In the design, we tried to be as ecologically valid as possible. In the image treatments, priors were not imposed rather they were simply the actual prior probabilities the subjects walked into the lab with. The image qualities, physical attractiveness and intelligence, are both economically and socially relevant. The ego utility or status of these qualities was not artificially imposed. Rather it relied on the subjects' underlying preferences. While many signals in the real world are of a subjective nature (compliments, praise in letter of recommendation) objective up/down comparisons are quite common as well.[7]

This section has 3 parts. We first present the analysis of the posterior beliefs, then we examine the round-to-round changes in greater details. Finally, we argue that the gender differences observed support our claim that self-image association is driving the results. 7 of the 150 subjects were eliminated from the analysis because they clearly displayed a lack of understanding of the experimental protocol.[8]

1.3.1 Posteriors

We label "up" messages $s = 1$ and "down" messages $s = 0$. Recall, $s = 1$ indicates the subject's rank is closer to 1 than the comparison participant. In *Beauty* rank 1 represents the subject whose physical attractiveness rated highest, in *IQ* rank 1 represents the subject with the highest IQ test score and in *Control* rank 1 represents the subject whose random "card number" (1-10) is 1.

The purpose of this section is to isolate how favorable ($s = 1$) and unfavorable ($s = 0$) news is incorporated into posterior beliefs. Our hypothesis is that for the image tasks, subjects will respond more to favorable news and less to unfavorable news. For the control, no difference is expected. For each subject in each round we calculated what their posteriors should be according

[7]Examining subjective signals is also possible with our experimental framework. Suppose we had elicited subjects' priors in *Beauty* before speed dating and then directly after. The feedback during the conversations is a subjective signal and is of interest to study, however we used objective signals to provide closed form solutions and leave the analysis of subjective signals to future work.

[8]The first reason for elimination was not having beliefs that added up to 100%. The second reason group of subjects eliminated appeared to interpret "up" as meaning closer to 10 (as opposed to closer to 1) but then at some point realized their mistake. This was a function of the somewhat confusing feature of the English language in that "higher" means better in terms of rank in in a distribution but means the opposite in terms of the absolute number. 2 subjects had a computer failure which lead to partial data loss.

to Bayes rule, given the priors they reported in round 0. We then took the expected rank under this distribution and called this value the "mean Bayesian belief".

Panels A, B and C of Figure 2 plot subjects' observed mean belief by the mean Bayesian belief for *Beauty*, *IQ* and *Control* respectively. The data is split by signal valence. It compares subjects who have received all good news to all bad news. This retains all observations from round 1 (when the subjects have only received 1 signal) and eliminates some observations from rounds 2 and 3; this is done to get the tightest comparison possible. The eliminated observations are analyzed in the following subsection and tell the same basic story.

The fitted lines are from the OLS regressions presented in Table 1. Optimistic updating occurs when the OLS fitted line is steeper for good news as compared to bad news. This demonstrates that subjects are reacting relatively more than the Bayesian for favorable signals and less than the Bayesian for unfavorable signals. A constant slope across signal valence indicates that bad news and good news are incorporated into beliefs symmetrically.

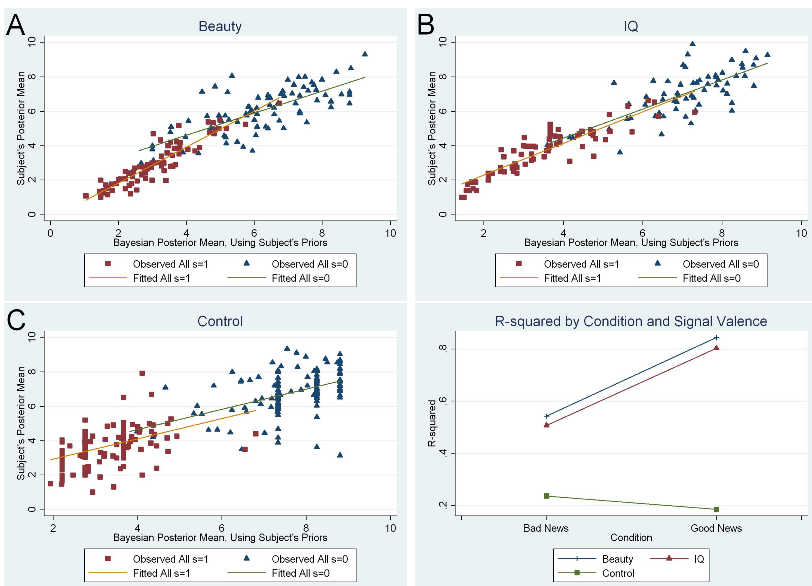


Figure 1.2: Posterior beliefs by signal valence and condition.

We see that the slope is indeed constant for in *Control* but that is steeper for good news in both image conditions. The effect is more pronounced in *Beauty* and table 1 shows that the difference is highly significant (as indicated by the coefficient on the interaction term). Subject's in *Beauty* engaged in optimistic updating as compared to the control. The difference in *IQ* is not statistically significant, however we see in panel D that subjects did differentially process information in this condition. In particular, subjects responded more predictably to good news, even if with less bias than in the *Beauty* treatment. Their response to bad news was noisier. The following subsection confirms this. The intuition, which is expanded upon in Section 5,

Table 1.1: Table 1: Subject's Mean Belief as a Function of Bayesian Mean

	Beauty	IQ	Control
Bayesian μ	0.641***	0.846***	0.589***
	(0.0620)	(0.0868)	(0.0862)
Bayesian μ*1{All $s = 1$}	0.406***	0.0714	-0.00118
	(0.0731)	(0.107)	(0.155)
1{All $s = 1$}	-2.311***	-0.580	-0.531
	(0.429)	(0.677)	(0.790)
Constant	2.054***	1.051	2.291***
	(0.412)	(0.642)	(0.666)
Observations	163	139	292
R^2	0.867	0.869	0.717

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

is that the pattern of updating discussed in the introduction has the property that whether or not net optimism emerges depends on prior beliefs. In *Beauty* subjects showed more initial overconfidence, this means that on average down signals were more informative than up signals. Not respecting signal strength for unfavorable news leads to more bias in this case.

Panels A and B provide visual evidence that beliefs are far more noisy for bad news as compared to good news for the image treatments. No such patterns emerge in panel C (*Control*) – the intercept shift displayed in Panel C is well within 1 s.e. of 0. Panel D presents the R^2's from the regressions in Table 1 separated by signal valence (i.e. 2 per condition, 1 for s=0 and 1 for s=1). The differences in R-squared are statistically significant (as given by a variance ratio test) for *Beauty* ($p = 0.0000$) and *IQ* ($p = 0.0000$) but not for *Control* ($p = 0.6139$) The result of the R^2 comparison is striking and is visually apparent in Panels A-C. . R^2 was 58% higher for good news in *IQ* and 60% for good news in *Beauty.* For the image treatments (where the notion of good vs. bad news makes sense) subjects' beliefs adhered far more tightly to Bayesian rationality when the news was favorable than when it was unfavorable. This is seen through a slope coefficient closer to 1 and much higher R^2. The control shows that this was not simply an artifact of the experimental application of ranks.

1.3.2 Round-to-round Changes

The preceding subsection examined how posterior beliefs compared to their Bayesian counterparts. In this subsection we examine how beliefs change from one round to the next upon the arrival of new information. Again, the analysis is designed to isolate differences in information processing by signal valence. Recall a negative change is a changes towards 1, the highest rank.

Figure 3 presents linear fits of the round-to-round changes in mean beliefs as a function of the Bayesian change.[9] The associated regressions are in Table 2. The inclusion of higher order terms does not significantly improve fit and the linear analysis simplifies the comparisons between treatments. Across all conditions and signals, there is a pattern of over-responding to uninformative signals, i.e. the y-intercepts are not zero. The importance of this result will be discussed in the following section.

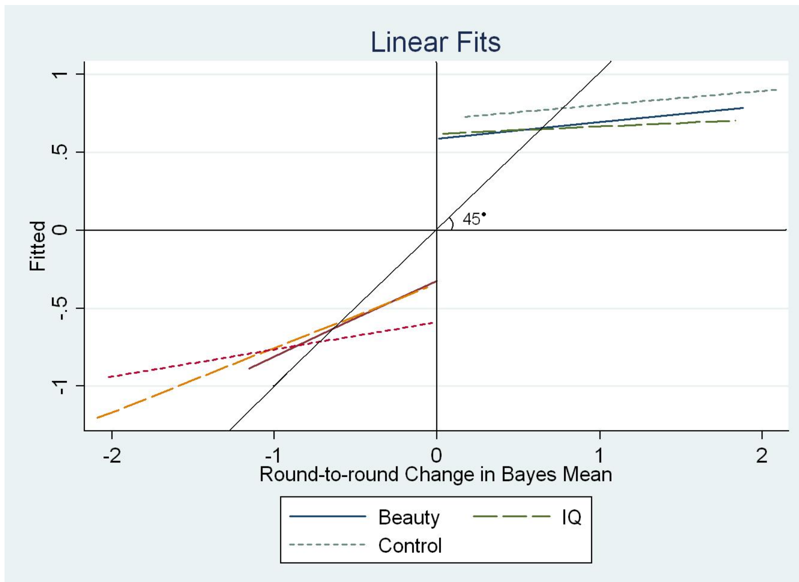


Figure 1.3: Round-to-round changes in mean belief by condition.

The 45 degree line is given for comparison to the Bayesian response. For *Beauty* and *IQ*, when the news is favorable (lower left quadrant) the fitted curve is much closer to the 45 degree line and has significantly steeper slope – signal strength is respected. The same does not hold true for unfavorable news. Table 2 shows that it is statistically indistinguishable from 0. Informative down signals are treated in a similar fashion to uninformative ones; beliefs are adjusted down in a noisily.[10] In the introduction we dubbed this pattern the good news-bad news effect. The effect shows up in both the posterior beliefs and the round-to-round changes.

[9]Both Figure 3 and Table 2 limit the analysis to changes in the correct direction. There were equal of number of changes in the wrong direction by signal valence. The inclusion of these outliers actually strengthens our argument as 2 were informative down signals in which the subject actually updated slightly updated towards 1. However, these points have such high leverage in the regression and represent less than 3% of the sample. As such they are excluded. Median regression is used in Table 2 to further reduce the influence of outliers.

[10]Our results would be considerably strengthened if we applied a kernel smoothing to prior beliefs. This is because Bayesian updating cannot put any weight on ranks that have prior probability equal to 0. For bad news, this tends to lead to an understatement of the optimistic updating because it was fairly common for a subject to have beliefs 50-40-10 on ranks 1, 2 and 3 respectively. The Bayesian posteriors have support of ranks 1-3, as such the mean does not move very much. However, it was exceedingly rare to see such concentrated (and likely mistaken) beliefs on the low end of the distribution. A smoothing procedure which transformed 50-40-10 to, for instance, 45-35-8-4-3-2-1-1-1 would lead to a much larger required downward change. In reporting beliefs, subjects generally left the possibility open that they were at the top of the distribution, so the smoothing procedure would not affect the analysis of up signals significantly.

Table 1.2: Mean Belief Changes by Signal Direction and Condition Dependent Variable: $\Delta\mu$

Condition	**Beauty**	**IQ**	**Control**
$\Delta\mu_{Bayes}$	0.212**	0.0256	0.141
	(0.0833)	(0.112)	(0.153)
$\Delta\mu_{Bayes} \times 1\{s=1\}$	0.475***	0.540***	0.141
	(0.165)	(0.149)	(0.205)
$1\{s=1\}$	-0.642***	-0.772***	-1.119***
	(0.107)	(0.134)	(0.288)
Constant	0.437***	0.599***	0.625***
	(0.0826)	(0.102)	(0.222)
Observations	206	183	385
R^2	0.49	0.55	0.48

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

For the image tasks, signal valence fundamentally affects how new information is incorporated into beliefs. There do appear to be similar patterns in *Control* albeit of a much lower magnitude and lacking statistical significance.

In the preceding subsection we showed that good news led to a much tighter adherence to Bayesian rationality than bad news. The following chart compares the standard deviation of the residuals for the regressions in Table 2, by signal valence. We can see the same pattern emerges, updating after down signals (s=0) is far more noisy in *Beauty* and *IQ*, but the essentially the same in *Control*. A variance ratio test gives $p = 0.0000$ for *Beauty*, $p = 0.085$ for *IQ* and $p = 0.347$ for *Control*. Once again the “image effects” in *Beauty* appear strong than *IQ* but are non-existent in *Control*.[11]

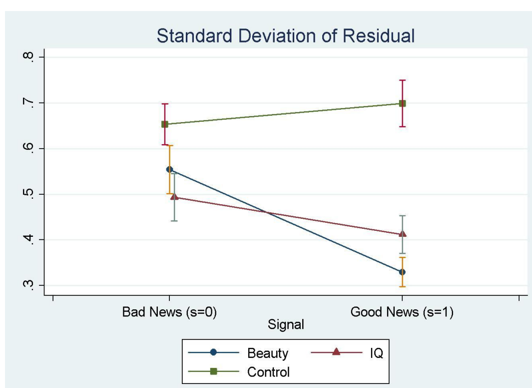


Figure 1.4: Noise in round-to-round updating by treatment and signal type

[11]It is worth noting that since social comparisons were used in *Beauty* and absolute comparisons in *IQ* we cannot concluded that physical attractiveness has a stronger image association than intelligence. We elaborate on this point in Section 3.3.

1.3.3 Image association and gender

We have argued that the chief difference between *Beauty*/*IQ* and *Control* is the degree to which the underlying quality is associated with self-image. *Control* uses random assignment of ranks in order to maintain image neutrality.[12] We have seen that differential processing of good versus bad new was most pronounced in *Beauty*. Under our working hypothesis, this would indicate physical attractiveness, as given by peer subjects in a speed dating exercise, has a greater impact on self-esteem for most subjects than intelligence, given by an IQ test.

There are many plausible reasons that this is the case. For one, formal feedback on physical attractiveness is not as common as formal feedback on intelligence. By the time they reach college, students receive many informative signals on intelligence such as SAT score, high school class rank and college GPA. There are no such analogs for appearance. [13] Rank out of 10 subjects on the appearance dimension is potentially more important to self-image than their rank on the intelligence dimension as the rarity of formal feedback on appearance adds to its ego-impact. Another reason is that the "beauty scores" were formulated by subjects in the room while the IQ score used an external standard. Many experiments have shown that subjects care about the opinions of their peers in experimental settings (see for instance Andreoni and Bernheim, 2009).

There is evidence that possessing physical attractiveness is more important for females and possessing intelligence is more important for males in attracting potential mates.[14] In Table 3 we look at how subjects beliefs changed relative to the Bayesian separated by signal history and gender. For both genders the values are mostly negative for *Beauty* as shown in Figure 2. We see that women were more optimistic in the face of the worst news (all $s = 0$, $t = -1.47$). In contrast, men were more optimistic in the face of the worst news in *IQ* ($t = 1.50$). The results indicate that men have a stronger aversion to believing themselves to be unintelligent, while women exhibit this aversion for physical unattractiveness. The effect only appears at the bottom the distribution. From a mate selection perspective this makes sense if finding mates becomes very difficult at a certain point on the left tail of the distribution.

1.4 Search Results

Our results in the last section showed that subjects dealt with the unwelcomeness of bad news by underweighting "down" signals. In this section we examine how people seek out

[12]It is possible that some people consider themselves lucky and associate a high rank with having good luck. Also, even the artificial application of rank could have an effect due to an innate responsiveness to relative comparisons. Indeed, there does appear to be a slight effect consistent with this reasoning.

[13]Subjects were asked in the post-questionnaire whether they had ever posted a picture on HotorNot.com or similar website. Only 4% of subjects answered that they had.

[14]The Trivers-Willard hypothesis posits that if a couple has traits particularly beneficial to one gender then their offspring should more likely be that gender [Trivers and Willard, 1973]. Using a large scale survey, Kanazawa (2006) finds that beautiful parents are significantly more likely to have daughters. [15] Based on these findings, it is natural to expect more optimistic updating in *Beauty* for females and *IQ* for males.

Table 1.3: Total Change in Mean Beliefs Relative to Bayesian $\mu_t - \mu_0$)-($\mu_t^B - \mu_0^B$) by Gender

	Round 1		Round 2		Round 3		All
Signal History							
Signal History	F	M	F	M	F	M	Pooled t test
Beauty Condition							
0 $s=1$	-0.13	0.14	-0.27	0.20	-0.47	0.02	-1.47
1 $s=1$	0.04	-0.12	-0.15	-0.07	-0.16	0.01	0.05
2 $s=1$			0.11	-0.19	-0.35	-0.31	0.90
3 $s=1$					0.18	-0.17	0.86
IQ Condition							
0 $s=1$	0.04	-0.28	0.04	-0.19	0.25	-0.44	1.50
1 $s=1$	0.44	0.22	0.11	-0.28	-0.41	-0.06	0.48
2 $s=1$			0.13	0.27	-0.09	0.01	-0.63
3 $s=1$					0.45	0.13	0.68

new information as a function of their current beliefs. Here signals are perfectly informative, so they cannot be selectively filtered. Our hypothesis is that the "sting of bad news" leads subjects to avoid learning their true rank as the probability that the revelation is unfavorable increases. Theoretical models have shown how overconfidence can develop and persist if agents acquisition of new information is a function of their prior beliefs. We will briefly discuss two such models and then analyze how subjects' WTP to learn their true rank varied by their final round posterior beliefs.

1.4.1 Theoretical Models

Santos-Pinto and Sobel (2005) presents a model in which image qualities are multi-dimensional . Agents can augment their skill and do so on the dimensions they believe are most important to the overall quality. While the model does not deal with information search directly, the underlying mechanism is one of egocentric search. In the real-world, people may also receive informational feedback during skill augmentation.

In Köszegi (2006) information acquisition affects utility in two distinct ways: 1) it allows for better decision making 2) it changes beliefs which enter utility directly. Information search thus depends on the payoffs to accurate beliefs and the shape of the ego utility function. Köszegi assumes that ego utility equals 1 when mean beliefs are above the 50th percentile and 0 otherwise. Information can be damaging to agents who currently believe themselves to be better than average – these agents will exhibit rational ignorance. The type of agents exhibiting rational ignorance depends directly on the ego utility function. We will see that a function with high utility of occupying the top 2 ranks and very low utility of occupying the bottom 2 ranks has good explanatory power in our setting.

Table 1.4: Willingness-to-pay as a Function of Final Round Beliefs

Condition	IQ	Beauty	Control
Final Round μ	-0.325***	-0.206**	0.0112
	(0.107)	(0.0855)	(0.0423)
Final Round σ	0.911**	0.917	0.237
	(0.382)	(0.575)	(0.174)
Constant	0.729	0.298	-0.339
	(0.523)	(0.750)	(0.336)
Observations	77	65	142
R^2	0.099	0.180	0.015

*** p<0.01, ** p<0.05, * p<0.1

Robust standard errors in parentheses

1.4.2 Results

As described in the methods section, after the 3 signal rounds subjects had an opportunity to learn their true rank. WTP is defined as the switching point on a price list offering a choice between receiving $\$x$ *and* learning their true rank on the task and receiving \$0 and *not* learning their rank. x ranged from -\$7.00 to \$7.00 in \$1 increments. As such subjects could require as much as \$7 subsidy to learn their true rank and could pay as much \$7. If the executed choice determined that a subject learn his true rank, then he had to initial by the rank on an "acknowledgment form."

Table 4 examines the relationship between WTP and the mean and standard deviation of their final round beliefs (i.e. after receiving all 3 signals). *Control* allows us to account for curiosity and other artifacts of the experimental setting unrelated to self-image.

Subjects WTP to learn their true rank increased as their round 3 posterior mean moved towards 1 (highest) for both *IQ* and *Beauty* – for *Control* there was not any effect.[16] In fact, subjects in the image conditions subjects who believed they were below average (rank 6+) often required a subsidy to learn their true rank. In the image conditions there was an "informational value" as well – WTP increased with the standard deviation of the belief distribution.[17]

If most of the ego utility effects are driven by the tails of the distribution then small sample sizes within each condition become a problem. This is why a linear fit was used in Table 4 despite its lack of realism. In Figure 5 we pool *IQ* and *Beauty* and analyze the data using a flexible non-linear form, fractional polynomials.

Figure 5 shows that subjects in the middle of the distribution have a flat WTP function – they are generally not willing to pay but do not require a subsidy to learn their true rank. The function is convex for high ranks and concave for low ranks. The very best subjects had a high

[16] There were not gender differences across condition.

[17] One might suspect that the informational effect might differ by mean belief. Regressions available from the authors show that there are not significant differences if the interaction terms by mean belief quartile are added.

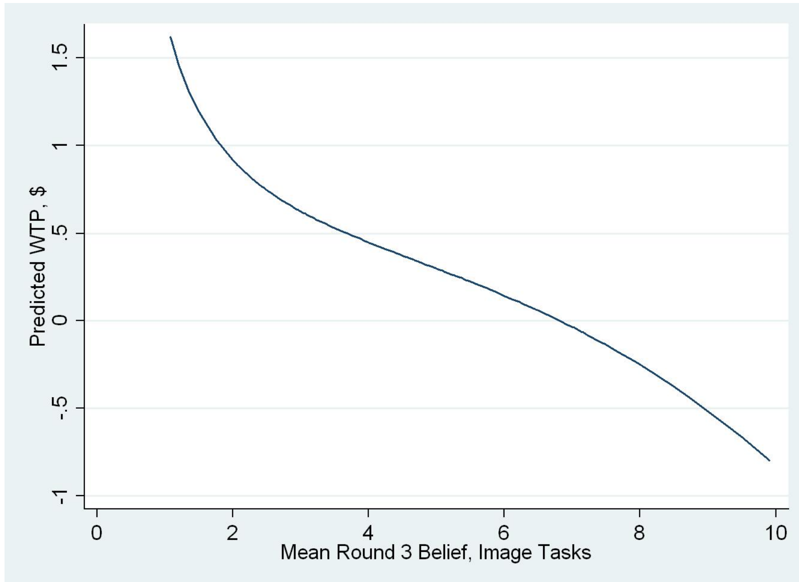


Figure 1.5: Fractional polynomial fit of WTP on mean final round belief in the image tasks.

WTP and the very worst required a subsidy.[18]

What type of ego utility function explains this data? Consider a subject with most of the probability mass of their beliefs in the top 3 ranks. In our data, on average this subject has positive WTP to learn their true rank. This is precisely the opposite behavior generated by Köszegi's functional form. In his model, the person would not want to risk receiving bad news. For this person to risk receiving bad news it must be that the potential ego gains outweigh the losses. This is the case if the ego utility function has increasing marginal utility as we move towards rank 1. There also must be increasing marginal disutility from median beliefs towards rank 10 so that subjects with the lowest round 3 beliefs require the largest subsidy to learn their true rank. Since agents in the center of the distribution have a flat WTP function, we infer ego utility is flat in this "average region." Essentially, subjects placed a high premium of knowing for certain they were the best in the room and a had a strong fear of learning they were the worst in the room.[19]

The type of search found has the property that beliefs will always be more precise as we move towards the top of distribution. A consequence is that high types are likely to know with precision their percentile rank while low types will have more diffuse beliefs – nobody will

[18] There is some evidence that male subjects have a higher WTP when they have beliefs near 1 as compared to females and females require a larger subsidy when beliefs are near 10. However, the differences are not significant. A larger sample size would be needed to examine gender differences at the tails of the distribution.

[19] An alternative plausible explanation, albeit one we view unlikely, is based in the psychological concept of "affirmation." Imagine an agent who is 99% sure she is rank 2. Suppose despite her near certainty, she has a high WTP to learn the truth. Positive (negative) feedback that has little informational value is known as affirmation (disaffirmation). Affirmation (disaffirmation) seeking (avoidance) has some explanatory power in our setting, however the significant effect of of round 3 belief variance shows that it is not the entire story.

believe with a high degree of confidence that they are in the bottom percentile ranks. This pattern of beliefs is not classic overconfidence (mean beliefs above average), but it can have economic consequences if people base decisions on the first moment of their distribution.[20] An example germane to our study is mate selection. Suppose nobody's mean belief places them in the bottom quartile of mate desirability. This can potentially disrupt the sorting necessary to reach equilibrium. Our second example involves a classic question in labor economics. Suppose the returns a college degree are positive only for those in the top 70% of the population. If subjects choose the mean of their beliefs to make decisions, then a lack of precision for below average types can lead them to enroll at sub-optimally high rates.

In round 0 subjects provided their beliefs with out an informational feedback. Our findings in this section are consistent with the prediction that below average subjects have a higher standard deviation in their beliefs than above average subjects. Across image tasks, the mean standard deviation of round 0 beliefs for above average types is 1.52, whereas it is 1.94 for below average types. The difference is highly significant. ($p = 0.00005, t = 4.263$). Overall, the precision heterogeneity in subjects' round 0 priors is consistent with the testable implication of our informational search findings.

1.5 Discussion

In this section we discuss how the good news-bad news effect can provide an explanation for observed confirmation bias and self-serving bias. We also illustrate how optimistic updating and differential search can compliment each other in generating overconfidence on image qualities, discuss the theoretical underpinnings of our findings and suggest real-world applications.

1.5.1 The good news-bad news effect and confirmatory bias

In Section 3 we showed that subjects incorporated favorable and unfavorable news into their beliefs in fundamentally different ways – the tighter adherence to Bayesian rationality for good news was striking. Another well documented bias involving differential processing of information is the confirmatory bias (CB). The classic account of CB is that people overreact or react appropriately to new information that agrees with their prior belief and always underreact to information that disagrees with their prior belief.[21] The precise definitions of "agree" and "disagree" can be murky in some settings. We think in our experiment the most natural clas-

[20]The effect on mean overconfidence depends on the signal informativeness, whether priors exhibit over/underconfidence and whether agents update rationally. For instance, if agents are overconfident and signals are perfectly informative, then this search will tend to reduce net overconfidence. If agents are rational and signals are imperfectly informative then this ego utility function will actually induce underconfidence. This results follows from proposition 1 of Köszegi (2006). We discuss how this type of search combined with the inference pattern we observed can lead to overconfidence in Section 5.3.

[21]Sometimes confirmatory bias is used to explain seeking information that is unlikely to be disconfirming (failing to test alternative hypotheses). This search-based CB finds support in our data.

Table 1.5: $\Delta\mu - \Delta\mu_{Bayes}$ as a Function of Prior Group and Signal Valence

	Good news		Bad news	
	Above avg	Below avg	Above avg	Below avg
Beauty	-0.14	-0.13	0.00	-0.13
	(-3.74)	(-1.20)	(-0.03)	(-1.32)
IQ	0.07	-0.08	-0.14	0.02
	(1.46)	(-0.70)	(-0.91)	(0.23)
Control	0.13	0.21	-0.36	-0.37
	(1.69)	(2.11)	(-5.16)	(-3.89)
CB prediction	$-$	$+$	$-$	$+$

t statistics ($H_0 = 0$) in parentheses

sification is as follows. For subjects whose mean belief places them in the top 5 ranks (above average) $s = 1$ is confirming and $s = 0$ is disconfirming. For subjects whose mean belief places them in the bottom 5 ranks (below average) $s = 1$ is disconfirming and $s = 0$ is confirming.

Results presented in Section 3 already indicated that our data show no evidence of CB. For instance, the lower responsiveness to bad news for subjects who believed they were below average, as shown in Figure 2 and Table 1, is precisely the opposite of the prediction of CB. In Table 5 we look for evidence of CB directly in comparing the subjects' round-to-round change in beliefs to the Bayesian change for confirming and disconfirming signals ($\Delta\mu - \Delta\mu_{Bayes}$). Negative values indicate optimistic updating towards rank 1.

Examining *Beauty* we see above average subjects did tend to overreact to good news, consistent with CB. However, it also evident that below average subjects receiving $s = 1$ overreacted with commensurate magnitude (0.01 less). Specifically they did not ignore or under-respond to the disconfirming good news. For $s = 0$ is it is a similar story, CB predicts that below average subjects would over-respond to $s = 0$ but the opposite occurs — there is an under-response. For IQ, the patterns are similar but of a lower magnitude. What appears to matter most in the image conditions is the signal valence, not whether a signal confirmed the prior belief. The patterns in the *Control* do not exhibit CB either. The negative values for bad news are also seen in the lower constant in the regressions in Table 3.

The question remains: what is the underlying cause CB? If it were a hardwired feature of human updating machinery, then we would expect to see it in updating on image tasks. Yet we do not. A CB agent has an aversion to switching their beliefs and admitting they were wrong. One reason could be that they get ego utility (self-esteem) from believing they are right.[22] Disconfirming signals are bad news while confirming signals are good news. When the underlying quality has direct self-esteem consequences, we do expect CB because the ego utility of being right is outweighed by the direct image effects. This is precisely what we observe in our

[22] It could also be anticipation utility. Suppose an agent made decisions today which pay out in the future based on the accuracy of their beliefs. If they change their mind, it means these actions will have a low expected payout and this lowers "anticipatory utility" today.

experiment. It also explains why we do not observe CB in *Control* – there is very low utility of being right about a random number. It is no surprise then that evidence of CB usually involves beliefs that do not have a direct image consequence but that are tied to identity (e.g. sides of a hot political debate [Plous, 1991, Lord et al., 1979] or beliefs which drove costly effort in the past [Mahoney, 1977]).

In past experiments showing CB, confirming signals were always good news. The perfect co-linearity of confirmation and valence rendered it impossible to distinguish between them. In our experiment, confirming or disconfirming news could be intrinsically favorable or unfavorable. In the image conditions what mattered was signal valence, not whether the signal confirmed the prior belief. In the context of the "psychological defense mechanism" discussed in the introduction, this result is not surprising. There is extensive evidence that people are more likely to remember good as opposed to bad events (Mischel, 1976), attribute good outcomes resulting from their actions to skill and bad outcomes to luck (attribution theory Heider, 1958) and shield themselves from uncomfortable truths such as the danger of their jobs (cognitive dissonance Feisteiger, 1957 and Akerlof and Dickens, 1982). The good news-bad news effect places CB within the realm of these self-protecting defenses.

1.5.2 Self-serving bias and signal valence

Self-serving bias is the tendency to conflate self-interest with moral judgment and can lead to bargaining impasse [Babcock and Loewenstein, 1997]. There is strong experimental evidence for self-serving bias [Messick and Sentis, 1979, Babcock et al., 1995]. In a typical experiment subjects are randomly assigned a role in bargaining game. The role involves a detailed back story and the subject's aim is to negotiate the best deal possible for themselves. Despite the fact that roles are randomly assigned, both groups tend to report that they deserve a greater share of the pie *based on the facts of the case.* Self-serving bias has also been shown to have economic and health consequences in the field [Larwood, 1978, Babcock et al., 1996].

In bargaining, signals that one will win the case are have positive valence – they indicate that a financial windfall is forthcoming. They might also mean that one did not act wrongly in the past as in, for instance, a divorce hearing or patent infringement case. The good news-bad news effect predicts that these positive signals will receive disproportionate weight as compared to negative signals, and consequently both sides will exhibit excessive confidence that they are in the right. Notice that CB has a tough time explaining self-serving bias, especially in controlled experimental settings. With random assignment, it is unreasonable to posit that ex-ante both groups of subjects believe they will be placed in the role that "should" win. Signals cannot be confirming or disconfirming in this setting. Conversely, since winning means higher earnings, news has intrinsic valence.

1.5.3 The joint effect of the search and inference results

In Section 4 we showed that WTP to learn true rank increased as beliefs moved toward rank 1. Our design used revelation of the truth (instead of just another up/down comparison) to amplify the differences in WTP and increase the power of our statistical tests. It useful to consider how the results would apply to situations in which signals are not perfectly informative. Figure 3 shows that people tend to over-respond to relatively uninformative signals and do not respect the strength of informative bad news. When one believes already believes they are high in the distribution good news is uninformative (it should be expected) while bad news is very informative. This means that if high types are seeking signals and generally receiving positive feedback, then they will develop very optimistic beliefs.

Biased search and processing can also lead to path dependence in beliefs. Path dependence in beliefs is stressed in Rabin and Shrag (1999)'s model of CB. Biased search and inference has a similar effect . The bias induced by the good news-bad news effect alone increases with initial overconfidence because bad news becomes increasingly informative. Our search findings suggest these are the signals likely to be actually received. Overconfidence on image tasks will develop when priors are accurate or optimistic, but it might not develop if beliefs evolve to (or begin at) a sufficiently underconfident state. In the latter case, underconfidence can get locked in as people avoid new information.

One application of this result is to the finding that people tend to be overconfident on easy tasks and underconfident on hard tasks (for a thorough review see Moore and Healy, 2008). If all comparisons are relative, then the distribution of good news and bad news will be independent of task difficulty. However, for absolute feedback (e.g. test score as opposed to percentile rank) bad news is more likely for a difficult task and good news for an easy task. If people over-respond to relatively uninformative success on an easy tasks, overconfidence will develop. This observation suggests that it would be nice to look at information search when people can choose between absolute and relative information. Our results suggest that they would choose relative comparisons for difficult tasks and absolute comparisons for easy tasks.

1.5.4 Explaining the good news-bad news effect

We have argued thus far that a direct concern for self-esteem and self-image endows signals with intrinsic valence, which leads to differential processing. The notion of a direct utility to beliefs is gaining purchase among economic theorists. Akerlof and Dickens (1982) first used belief-based utility in their model of cognitive dissonance.[23] In their model, agents rationally maintain optimistic views on the danger of their jobs in order to avoid "fear utility" in the interim. Benabou and Tirole (2002) and Brunnermeier and Parker (2005) generalize this notion

[23] The notion of cognitive dissonance was introduced formally by Festinger (1957). For a review see Simons et al. (1970).

calling it "anticipation utility" and show that a rational agent will always "choose" an updating rule that is biased towards optimism. Compte and Postlewaite (2004) modify preferences so that self-confidence enhances performance and establish a similar result. The intuition behind these models is that using Bayes rule maximizes belief accuracy, so moving away from Bayes rule toward optimistic updating has only a second order cost (through the loss in accuracy) but a first order benefit through either anticipatory utility or enhanced performance. As such a rational agent always prefers to induce some optimistic bias.

When beliefs affect utility in these "non-standard" ways, agents have an incentive to sacrifice belief accuracy in order to increase the belief based utility, blurring the distinction between hopes and beliefs. Normatively speaking, the sacrifice in accuracy should depend on the marginal returns to belief accuracy as compared to the marginal returns to belief utility. Whether the good news-bad news effect is rational or not depends on whether agents can overcome the tendency to discount bad news when the returns to belief accuracy increase. An experimental design which increases the costs to inaccurate beliefs while holding belief utility constant would be able to answer this interesting question. We leave this to future work.

1.5.5 Real-world applications

In our experiment signal valence was generated by ranking 10 strangers after a short task measuring their intelligence or beauty. In an experimental setting, ethical considerations rightly limit "how bad" news can be. Rank on the image task, as compared to 9 strangers, endowed the signals with enough psychological valence to significantly change their updating behavior. In the real-world, news can range from the overwhelmingly positive to the jarringly negative. In these situations we would expect differential processing by signal valence to play an even larger role than it did in the experimental setting. We have already discussed real-world applications in which the effect of valence is mediated through CB, overconfidence and self-serving bias. However, there are applications that fall outside the scope of these previously identified inference mistakes. We discuss a few such examples below.

In finance, the good news-bad news effect predicts that an agent will respond differently to news on a security she owns as compared to news on a security that she does not own. If people have a difficult time responding objectively to news on securities they own, and on some level recognize this, then it make sense to hire financial professionals to do the job for them. They are willing to pay a financial manager for doing a job that an educated person could quite easily accomplish, at a far lower cost, by purchasing mutual funds through a discount online trading account.

Further evidence from finance come from the literature on post-earnings announcement drift (PEAD). Predictable drift in stock prices should not occur in an efficient market, but many authors have found that stocks experiencing earnings surprises do predictably drift over a 120 day period following earnings surprises [Foster et al., 1984]. Bernard and Thomas (1989) found that the drift is significantly more pronounced for bad news stocks as compared to good news stocks. That is, the initial price movement accounts for less of the available information for bad news as compared to good news, which is precisely our primary finding. The authors show that these patterns are driven by small-cap stocks (those traded by non-institutional investors). For large-cap stocks good news and bad news are treated the same. The authors also rule out risk-premium-based and other explanations for the observed drift.

An application in medicine is a hospital patient's prediction of their prognosis – here news is literally life-or-death. Studies have found that cancer and HIV+ patients optimistically assess their chances of survival [Weeks et al., 1998, Eidinger and Schapira, 1984]. It is not, however, that their physicians are misleading or giving them intentionally optimistic forecasts; the physicians' forecasts are shown to be accurate and unbiased. In fact, the phenomenon has attracted attention in the medical community because it complicates physicians' ability apply treatment they find most appropriate given the circumstances. An optimistic patient is liable to make decisions that are not in her best interest and physicians have adopted strategies to try to alleviate this problem [Paling, 2003]. A related application is to preventative care. Studies show that Americans tend not to heed early warnings signs of disease, which drives up health care costs in the long-run [Schuster et al., 2005]. Early signs of illness are unwelcome bad news and our prediction is that this news is discounted.

1.6 Conclusion

We have shown that image effects have an important implication for Bayesian inference. When updating on a attributes closely tied to self-esteem, new information is not only confirming or disconfirming but also intrinsically favorable or unfavorable. In our image treatments, good news and bad news were incorporated into prior beliefs in entirely different ways. For good news, the inference is less noisy, respects signal strength and conforms much more closely to Bayesian rationality. Conversely for bad news, beliefs are more noisy and scarcely resemble the normative Bayesian standard.

The image treatments used attributes designed to have a strong self-esteem association: rank out of 10 subjects in the session on intelligence as given by an IQ test and physical attractiveness (rankings within gender group) as given by peer ratings in a speed dating exercise. In our control condition we used randomly generated subject ranks. Here mistakes were made, but they were largely symmetric with respect to signal direction. In further support of our main finding, gender differences in image effects were consistent with studies in the mate selection literature that show physical attractiveness has a higher importance for women and intelligence a higher importance for men.

In Section 4 we further examined the "sting of bad news" by providing subjects with the opportunity to learn their true rank on both the image task and control. Subjects who believed they were near the top of the distribution had a high WTP to learn their true rank. Conversely, subjects who believed they were near the bottom of the distribution required a subsidy to learn the truth.

Past work has shown that humans have a robust "psychological defense mechanism" to maintain positive self-esteem. A separate literature has examined systematic departures from Bayesian rationality in signal inference. This paper draws from both literatures to study how people incorporate objective signals into their existing beliefs as the degree of self-esteem impact is varied. The results point to differential processing by signal valence as the underlying cause of confirmatory bias, self-serving bias and overconfidence. Furthermore they place these biases within the realm of the psychological defense mechanism and confirm the predictions of theoretical models that predict valence-dependent updating because agents have a direct utility to beliefs. The paper suggests that future research could benefit from further synthesis of these two fields.

1.7 Appendix

1.7.1 Experimental Instructions

Instructions taken from a *Beauty* session.

Welcome

Thank your for participating in our experiment. We will begin shortly. Today's experiment will last under an hour and half.

Informed Consent

Placed in front of you is an informed consent form to protect your rights as a subject. Please read and sign it. If you would like to choose not to participate in the study you are free to leave at this point. If you have any questions, we can address those now. We will now pass through the aisle to pick up the forms.

Anonymity

Your anonymity in this study is assured. Your name will never be collected or connected to any decision you make here today. Your email was collected for invitation purposes only, it will never be connected to your performance in the study. Furthermore, your earnings will be paid in a sealed envelope with your subject letter ID so that even those running the study will not know your earnings. No other subject in the study will know any of your choices or performance in the study.

Rules

- Please turn your cell phones off.
- If you have a question at any point, just raise your hand.
- Please put away any books, papers, computers, etc that you have brought with you.

Stages of Today's Experiment

Today's experiment will have two stages. The instructions for each stage will be read at the beginning of that stage. At the conclusion of stage 2, your payments will be prepared and the experiment will end.

Your Earnings

Some of the decisions you make today will impact your earnings. We will explain exactly which decisions and how earnings will be calculated at the appropriate time. Your earnings will be paid in cash, placed in a sealed envelope with your subject number. Although earnings will vary by subject, most subjects will earn between $20 and $30.

Stage 1 Instructions, Part 1

You will notice that this experiment is exactly half males (Subject ID M-A through M-J) and half females (W-A through W-J). Shortly we will begin a "speed dating" exercise in which you will meet 5 members of the opposite sex for 4 minutes a piece. Suggested conversation topics have been provided for you. Try to stay on these topics. These are:

- What are your plans following college?
- What is your favorite book/movie/album?

Following each 4 minute "meeting" you will fill out a brief meeting questionnaire. On this you will rate the person you just spoke with on three categories: ambition, attractiveness, friendliness. This will be done at a separate work station so that no other subject will ever see your ratings.

In your subject packet are is meeting questionnaires with your subject ID and the subject ID letters of the 5 participants you will meet. Males will move clockwise around the room on the switch. You will fill out the questionnaires privately at a different station. Under absolutely no circumstances will any other subject see the ratings you record. Your honesty is paramount to the aims of this study. After you fill out the form, place it in the manila envelope with your subject letter and proceed to the next station. We will now ask all the males on the inner column to switch with the females across the aisle and show you exactly how this stage will proceed.

Stage 1 Instructions, cont.

Please return to your original seat and we will proceed with the study.

We are now passing you a sealed envelope with your subject letter ID written on it. Within each is a note card with a number on it. For the males, these numbers range from 1-10, with each male having 1 number. The same applies to the females. For the remainder of the study, we will refer to the left side of the room as the "male group" and right side as the "female group." Within each group, each one of you has different number. The numbers have been assigned randomly by rolling a ten sided die prior to the study [subjects are shown the die]. The role of this envelope in the study will be explained at the appropriate time. You will not be allowed to open the envelope until the experiment has concluded. [stop] Are there any questions?

Things To Remember

The numbers in the envelope labeled "Card Number" range from 1 to 10, each subject in a given group has a different number. These numbers were assigned randomly. *Do not open the envelope.*

This Concludes Stage 1.

Stage 2 Instructions , Set 1 (Appearance Task)

Stage 2 consists of 2 "sets" each consisting of 4 "rounds."

In stage 2 we will ask you tell us the probability you think you occupy each of 10 ranks for the "task." The reason that there are only 10 ranks is that you will be making comparisons only to members of your gender group. So since there are 10 subjects in each group, this means that for each task there exists a ranking 1-10.

For this task, the appearance task, your rank is given by your "appearance score" which is the average of the 5 ratings on "appearance" in stage 1 on the experiment. The subject in your gender group with the highest average score is ranked 1, the one with the lowest, with your group, is ranked 10. A sample decision screen is now being displayed. You can control each "rank box" by clicking the slider and/or the arrows adjacent to the box. (Short description and demonstration).

In each round, you will be asked to fill in these boxes with what you think are the probabilities that your appearance rank is the number beside the box. You can do so by moving the slider or clicking the arrows. Notice that the card numbers give ranks 1-10, just as the boxes. These probabilities must add up to 100%. If they do not add up to 100%, you will be paid nothing for the round. The total is displayed to you on the screen. If they do add up to 100%, you will be paid according to the following formula:

$$2 - \sum_{i=1}^{1} 0(1\{rank = i\} - p_i)^2 \tag{1.1}$$

While this payoff formula may look complicated, what it means for you is simple: **you get paid the most on average when you honestly report your best guesses of the probability for each rank.** The range of payoff is 0-2 dollars for each round of guesses.

Each "decision form" is labeled by round and task number. These labels are on the tabs of the spreadsheet. (short demonstration moving between tabs). Once you complete a decision form, please click forward to the next tab. You may not go back and change your previous round guesses. However the guesses from your all previous rounds of that set will be displayed at the top of the screen. The spreadsheet records the time of entry and will check to make sure that all entries were down at the right time. You will not be eligible for payments if you have an entry that is out of order (i.e. if you went back and changed a previous entry).

Before rounds 1, 2, and 3, your rank will be compared to a randomly chosen "comparison participant" FROM YOUR GENDER GROUP. You will never know the identity of these participants. In each round, it will be a different participant. To convey the result of the comparison of ranks, we will pass you a "signal card" with the words "You are ranked higher" or "You are ranked lower" written on it. If you get the message "You are ranked higher" in a given round, it means that your appearance score was higher (equivalently your rank was closer to 1) than your comparison participant for that round. If you get the message "You are ranked lower" in a given round, it means that your appearance score was lower than your comparison participant for that round. For example, if your appearance rank was 5 and we compared you to a subject whose score ranked them 7, then you would receive the message "You are ranked higher." At the beginning of the next round, you will make a new set of guesses.

After receiving your message in rounds 1-3, click the box "Enter message here" and change it to reflect the message you actually received. This way we know you got the message.

In rounds 0-2, we will ask you to the probability you expect to get a "higher" message in the next round. Your payment for that guess will be made in the same fashion as the rest, so once again you have the incentive to truthfully tell us the probability you expect to to receive a "higher" signal in the subsequent round.

Things to Remember

- In round 0 you will make your guesses before receiving any messages.
- All rank comparisons and guesses are WITHIN GENDER GROUP.
- Your appearance score gives your rank for this task, highest score is rank 1 and so forth. The "signal cards" compare your rank to a randomly chosen "comparison participant" from your gender group.
- Before rounds 1, 2 and 3 we will compare your rank to a randomly chosen "comparison participant," which will differ by round.
- You will receive a message that says either "You are ranked higher" or "You are ranked lower"
- After receiving the message, you will make new guesses for the next round.
- In rounds 0-2, you will also make a guess at the probability of receiving a "higher" message in the subsequent round.
- You will make the most money on average by making your probabilistic guesses honestly.
- You will make your guesses by clicking the arrows or moving the slider adjacent to the "rank boxes" on the spreadsheet in front of you. The graph will change as you click the arrows or move the slider and the chosen percent for that rank will be displayed.
- Your guesses must total 100% in each round to be eligible for payment.
- Once you complete the decision form for a round, please click forward to the next round using the tabs at the bottom of the screen. You may not click back to change your answers as this will violate the aims of the experiment and make you ineligible for payments.
- Once everyone is done with the round, we will proceed to the next, so take your time when making decisions.

Stage 2, Set 2 (Card Number Task)
We will now begin set 2.
In the enveloped with your subject ID letter is a note card with a number from 1-10 on it. Within each gender group, one person has each one of these numbers.

This randomly assigned number gives your rank for this task, 1-10. Each of the 10 ranks is filled by exactly 1 subject within your group. Set 2 will proceed exactly like set 1, except that your rank is determined by your card number.

Things to Remember

- In round 0 you will make your guesses before receiving any messages.
- All rank comparisons and guesses are WITHIN YOUR GENDER GROUP.
- The subject whose card number is "1" is ranked 1. The "signal cards" compare your rank to a randomly chosen "comparison participant" from your gender group.
- Before rounds 1, 2 and 3 we will compare your rank to a randomly chosen "comparison participant," which will differ by round.
- You will receive a message that says either "You are ranked higher" or "You are ranked lower."
- After receiving the message, you will make new guesses for the next round.
- In rounds 0-2, you will also make a guess at the probability of receiving a "higher" message in the subsequent round.
- You will make the most money on average by making your probabilistic guesses honestly.
- You will make your guesses by clicking the arrows or moving the slider adjacent to the "rank boxes" on the spreadsheet in front of you. The graph will change as
- you click the arrows or move the slider and the chosen percent for that rank will be displayed.
- Your guesses must total 100% in each round to be eligible for payment.
- Once you complete the decision form for a round, please click forward to the next round using the tabs at the bottom of the screen. You may not click back to change your answers as this will violate the aims of the experiment and make you ineligible for payments.
- Once everyone is done with the round, we will proceed to the next, so take your time when making decisions.

Stage 3, Final Decisions

You will now have the opportunity to learn your true rank on the Appearance and Card Number task. Whether we reveal your true rank on the Appearance and/or Card Number task will be determined by your choices in the following decisions. The decision will proceed as follows. We are now passing out two "final decision forms." Each form is labeled either "Appearance task" or "Card Number task." Each form asks you to make 15 decisions, indicating whether your prefer option A or option B. Here is a sample of the decision form for the Appearance task: [sample shown]

You may only switch once from column A to B, if you switch at all. Why should you tell the truth? Once we collect the forms we will randomly choose a number (one for each form) between 1 and 15 inclusive and execute your decision for that number. Your final earnings will reflect your choices.

Example 1: Suppose you switch from A to B at decision (3). You are indicating that for a payment of $5.00 you'd be willing to learn the true rank on the task, but for $4.00 you would rather not learn the rank and receive $0. Suppose in this case we randomly draw 15. In this case you would not learn your true rank and receive $0. Suppose instead we draw 2. In this case you would receive $6.00 and learn your rank on the task.

Example 2: Suppose you do not care if you learn your true rank or not. In this case you are best served switching at 8. The reason is this maximizes your chance of receiving money (if 1-7 are drawn) and ensures you will never pay money.

What happens if it is determined you will learn your true rank?

After we collect your decision forms and determine the decisions that will be executed, we will know which subjects have chosen to learn their true rank. If it is determined that you will learn your true rank on either (or both) task(s), we will pass you a piece of paper with the rank(s) written on them. If it was determined that you will not learn your true rank, then this space will be left blank for you. For each rank that was revealed to you (if applicable) you must write the phrase "acknowledged, rank #" right next to the rank, where # is your rank on that task. We will then pick up the forms, making sure that everyone whose rank was revealed to them has acknowledged. We will pass out these forms to every subject, regardless of whether you chose to learn your true rank. As such, no other subject will know whether or not you learned your true rank. Furthermore, no other subject will ever learn your true rank. This process ensures your anonymity. We will pick up your envelope at this time as well.

How are any payments handled?

Your final earnings will reflect the monetary component of the executed decisions.

Things to Remember

- Your choices in this decision (and chance) will determine whether or not you learn you true rank on the Card Number and/or Appearance task.
- You can ensure that you always learn your true rank on a given task by always checking column A. You can ensure you never learn your true rank by always checking column B.
- If you do switch between column A and B, you can only do so once.
- Following these choices, the experiment will conclude.

- Two random numbers (1-15) will be selected to determine which decision will be executed for each form.
- If your choice indicates you learn your true rank you will be notified by an acknowledgment form, which you will initial to indicate you saw the rank.
- You may fill out the two forms in order you wish.
- You have a financial incentive to answer truthfully.

ACKNOWLEDGEMENTS

This paper was co-authored with my colleague from UC San Diego, David Eil. I would like to thank him for his contribution and support.

Chapter 2

“He Got Game” Theory: Optimal Decision Making and the NBA

Abstract

The paper uses standard theoretical concepts to examine the optimality of the shooting decisions of National Basketball Association (NBA) players using a rich data set of shots, collected by watching 60 games and recording the shot conditions and outcomes. Shot timing is modeled as an optimal stopping problem and the prediction of a monotonically declining “reservation expected shot value” with time remaining on the shot clock is confirmed in the data. For shot allocation, the mixed strategy Nash equilibrium predicts that each of the 5 players on the court offer equal expected point values per shot attempt. This prediction finds strong support in the data, especially among the starting line-up. Starters’ point values per shot were within a 0.75% band of each other (1 s.e.). Consistent with the model’s prediction, the more talented players (as measured by salary and awards) shot with a far higher frequency. In support of the model’s mechanics, star player Kobe Bryant is shown to significantly increase his teammates’ performance.

2.1 Introduction

A fundamental concern of economics is optimality: do agents and firms maximize in accordance with the normative predictions our theoretical models? This paper examines the problem of allocation of resources in order to maximize firm output - a classic line of inquiry, which is taken to a new testing ground. The firm being studied is a National Basketball Association (NBA) team and the agents making the allocation decisions are the professional players (and coaches). In many ways, the input allocation decisions facing the team are similar to those facing a "textbook firm" of classical producer theory. Both must allocate labor when productivity differs across workers and make decisions over time to maximize firm value in expectation. This paper examines the team's allocation of shots across players with differing ability and the timing of shots over the course of an offensive possession.

Given the nature of the economic activity engaged in by the firm being studied, the theoretical models developed to apply the optimality litmus test are rooted in game theory. Researchers have tested the predictive power of game theoretic equilibrium concepts in both the laboratory and field. The advantage of the laboratory is that the experiment can be tailored to test the falsifiable implications of the model, while the advantage of the field is that the agents (or firms) under study are performing a familiar task in a natural setting. Indeed the "poor" performance of experimental subjects has led to the use of experts with relevant experience as subjects.[1] The observability of outcomes in professional basketball offer the best of both of both worlds: experts operating in a natural setting with direct tests of the optimality of their actions.

Professional sports offer an ideal testing ground for economic theory – the stakes are high, the payoffs are often readily quantifiable and agents' actions are observable. Three recent papers in the spirit of this one have tested the predictions of theoretical models using the performance of professional athletes.[2] Walker and Wooders (2000) study serve location in professional tennis. Chiappori, et. al (2002) and Palacios-Huerta (2003) study penalty kicks in professional soccer.[3] While these papers differ substantially in their methodology, the upshot of all three is that mixed strategy Nash equilibria find strong support in the data. In the shot allocation problem studied in this paper, the *team* must mix over five pure strategies because there are five players on the court at one time, as compared to an *individual* mixing over two pure strategies as in each of the above cited papers. An additional contribution is the introduction of a new theoretical standard, the optimal stopping rule used to model shot timing, to the applied game theory literature.

In the game of basketball, players face shooting opportunities with variable defensive pressure and distance; further complicating matters is the fact that the shot must be taken before the 24 second shot clock expires. The team consists of players with differing shot productivity and must allocate shots efficiently across players to maximize performance. The paper focuses on two key aspects of shooting behavior: shot selection as a function of time remaining on the shot clock and allocation of shooting resources across players. A rich data set I collected allows the shot allocation and optimal stopping model to be applied to observed shooting behavior from the 2007/08 Los Angeles Lakers season. The data include distance, defensive pressure, shot type, shot fouled, location, shot blocked, shot fouled and time remaining the shot clock.

Shot timing is modeled using a finite period optimal stopping model in which players

[1] Two recent examples are [Palacios-Huerta and Volij, 2008] which examines chess masters play in the centipede game and [Fehr and List, 2004] which looks at play of CEOs in the trust game, with sanctions.

[2] Another example of using the easily quantifiable payoffs is [Massey and Thaler, 2007] who study of the loser's curse in NFL draft picks.

[3] The papers differ in that Chiappori, et. al aggregates across players while Palacios-Huerta uses a panel of shooters and goalies to study individual players.

face shot opportunities of varying expected point values and must decide between shooting and continuing. Since a shot must be taken in a 24 second period, the process has finite opportunities which potentially vary by possession. In this framework, the problem is similar to textbook “secretary” or “car buying” problems, which have been extensively studied in the lab. For example, Lee (2006) presents subjects with a sequence of up to 6 draws from a known distribution. A subject can stop the sequence at any point earning the value of the draw. He finds support for a monotonically declining stopping threshold. The observability of shot difficulty and time remaining on the shot clock allow me to extend these studies to the field.[4] Consistent with the model, the data provide support for a monotonically declining “reservation shot value” as the shot clock winds down to zero.

To examine shot allocation across players I construct a parsimonious model, which provides testable game theoretic implications of optimal shot selection. The model predicts that all players on the court at a given point in time should have the same expected point value per shot attempt. That is, the team should not be able to reallocate shots between players to improve the value of each possession and hence, it is argued, the chances of winning the game. This is a familiar feature of mixed strategy Nash equilibrium.

The prediction finds strong support in the data especially amongst players in the starting line-up (all of whom offer expected point values per shot that do not differ significantly from each other). In fact, when the analysis is limited to 5-man lineups which include the team’s 3 core players the resulting expected point values are even more tightly clustered around the “core average” of 1.37 points (within 0.75% for the starting lineup, who account for a strong majority of the team’s shots). There are 2 players who have statistically significantly lower performance (about 10%) than the 6 other players in the sample. These 2 back-up players appear to shoot too often, their attempts should have been limited in order to allocate the ball to their more productive teammates. However their shots account for a small fraction of the team’s total shots. One of these players was not offered a contract at the end of the season and the other experienced declining playing time throughout the season.

The results are robust to changes in the line-up. For instance, when 11-time All-Star and 2008 NBA Most Valuable Player (MVP), Kobe Bryant, is playing his teammates are more productive than they are when he is on the bench (i.e. he “makes his teammates better”) but they effect is uniform across players. This is not only consistent with the model’s equilibrium but also supports the underlying mechanics. In the model, when one player shoots more frequently, the defense responds by reallocating defensive pressure to that player. The reallocation creates better shooting opportunities for teammates, which is exactly what happens when Bryant is on the court.

A related implication of the model’s mixed strategy equilibrium is that the better players

[4] Authors have studied the stopping behavior of firms in harvesting tree stands [Provencher, 1997], renewing patents [Pakes, 1986] and replacing bus engines [Rust, 1987].

shoot more frequently but have the same expected point value per shot attempt as their teammates. This prediction is confirmed by the data. For instance, Bryant takes over a quarter of the teams shots but does not offer better value per shot than other starters. The second highest paid player on the team and United States Olympian, Lamar Odom, took the second most shots. The model asserts that a measure of a player is the ability to take many shots while maintaining a high productivity per shot and the capacity to increase teammates' productivity by attracting defensive attention away from them and not simply average shot value or total points scored. The empirical findings strongly support this assertion: point value is uniform across players but the more talented players bear a larger portion of the offensive burden.

The paper proceeds as follows. Section 2 describes the data. Section 3 models and analyzes shot selection as a function of the shot clock. Section 4 does the same for shot allocation across players. A conclusion follows in Section 5.

2.2 Data

The data used in the paper are 60 games from the Los Angeles Lakers 2007-08 National Basketball Association season. For each game, shot time, distance and outcome were first entered using the game logs (an online play-by-play including shot time and success/failure) available at espn.com and the NBA's official website.[5] Each game used in the study was recorded and the games were then watched by the author to record shot location, defensive pressure, touches for each player on the court, shot blocked/fouled, shot type and time on the shot clock.[6] The author performed all the game cataloging and did so in as objective a manner as possible.[7]

Table 1 describes the variables collected.

The Lakers were chosen because at the beginning of the season they were projected to be one of the better teams in the league and the fact that their star player Kobe Bryant is considered by many to be the league's top player.[9] Data was collected only for the Lakers and not their opponents. The NBA season is 82 games long, given the lengthy nature of the cataloguing process, it was determined that 60 games provided a large enough sample. Appendix 1 further describes the collection process.

The games for which the time on the shot clock was recorded produced 1,652 shots, the

[5] An example can be found at http://sports.espn.go.com/nba/playbyplay?gameId=280106013.

[6] For the first 34 games in the sample a binary variable was recorded when the shot clock was less than 5, for the remaining games the actual time was recorded.

[7] Many games have been transferred to .mpeg format and can be provided along with the coded spreadsheets to any parties interested in examining my cataloging system.

[8] A shot is counted as defended if the defender is within 2ft of the shooter and actively guarding him when the shot is attempted, notably this excludes "close-outs" where a defender runs at an open shooter. The defended variable was the hardest to code, particularly difficult cases were watched many times. In these cases the motion of the shooter was examined in an effort to determine if the nearby defender affected the shot. Every effort was made to apply as objective a standard as the other measures used in the study.

[9] These predictions turned out to be correct, the Lakers reached the NBA Finals and Bryant was named the league's MVP.

Table 2.1: Descriptions of the variables

Variable	Description
Shot Time	Time of the FG attempt, includes attempts when the player was fouled
Defended[8]	=1 if player was guarded at the time of the shot
Hit	=1 if shot is made
Double team	=1 if player was defended by 2 or more opponents
Touches	Number of touches on a possession, by player
Team touches	Total touches on a possession
Shot Zone	One of 14 zones giving physical location on the court
Turnaround	Fadeaway, turn-around jumper or hook shot
Distance	Shot distance in feet, =0 for dunks and 1 for layups
Pulled	Player removed from game
Fouled	=1 if shooter was fouled
Blocked	=1 if shot is blocked
Forced	=1 if shot is taken with less than 5 seconds on the shot clock
Shot clock	time on the shot clock
Offensive rebound	=1 if the shot resulted in an offensive rebound

complete data set comprises 4,522 shots. The key outcome variable for a shot is the expected point value, which are formulated as follows. First, shot success is modeled separately for each player using Probit with the variables listed in Table 1 as regressors along with a cubic in distance. Second, the probability of an offensive rebound (offensive team retains possession of the ball following a shot) is regressed on the same explanatory variables and player fixed effects. These regressions can be found in the appendix. The predicted success probabilities are converted to expected points for player i and shot t, E_{it} as follows:

$$\mathrm{E}_{it} = \begin{cases} \hat{p}_{it} * 2 + \hat{p}^{\text{off}}_{it}(1 - \hat{p}_{it}) * \mathrm{EV}_{t+1} & \text{fouled=0, three=0} \\ \hat{p}_{it} * 3 + \hat{p}^{\text{off}}_{it}(1 - \hat{p}_{it}) * \mathrm{EV}_{t+1} & \text{fouled=0, three=1} \\ \hat{p}_{it}(2 + f_i * 1) + (1 - \hat{p}_{it}) * f_i * 2 & \text{fouled=1, three=0} \\ \hat{p}_{it}(3 + f_i * 1) + (1 - \hat{p}_{it}) * f_i * 3 & \text{fouled=1, three=1} \end{cases}$$

where f_i is player i's free throw shooting success rate for the 2007/2008 season, EV_{t+1} is the expected value of a possession following an offensive respond, $\hat{p}$ is the predict shot success rate and $\hat{p}^{\text{off}}$ is the predicted probability of an offensive rebound for a given shot.

2.3 Shot timing as an optimal stopping problem

This section examines shot selection as function of time remaining on the 24 second shot clock for a given possession. By rule, if the shot clock expires the offense turns the ball over to the defense. This feature of game play allows the problem of when to shoot to be modeled as a finite period stopping problem.

2.3.1 Model

Each possession consists of N independent shooting opportunities for the team which, for the purposes of the model, acts as a single entity.[10] An opportunity is defined as a draw from the shot (expected) point value distribution, $p \sim N(\mu, \sigma^2)$. If a player decides to shoot, the possession ends and the team earns p (i.e. U(p)=p). Note that this implies the players perfectly observe shot the value of a shot opportunity. Given risk-neutrality this assumption is innocuous as tacking on a zero mean error in perception does not alter the decision problem ex-ante. In all periods before period N the team can decide to continue and receive more draws from the p distribution.

The solution to this problem is a series of cut points or reservation point values dictating a player shoot if the observed value p is above and continue if p is below the period cut point. It is solved by starting at the final period and working backwards through the opportunity sequence. In the last period the team must shoot and in expectation earns $\mu = \text{cut}_N$. In period $N-1$ the team must decide to either shoot or continue. Continuing provides expected utility of μ, as such the team will not shoot unless that draw $\tilde{p}_{N-1} > \mu$ which occurs with probability $1 - \Phi(\frac{\text{cut}_N - \mu}{\sigma})$ where Φ is the CDF of a standard normal distribution. If the team does in fact shoot at $N-1$ it earns $\text{E}(p|p > \mu)$. Putting it all together the expected utility of reaching period $N-1$ is given by:

$$\text{cut}_{N-1} = \left(1 - \Phi\left(\frac{\text{cut}_N - \mu}{\sigma}\right)\right) * \text{E}(p|p > \text{cut}_N) + \Phi\left(\frac{\text{cut}_N - \mu}{\sigma}\right) * \text{cut}_N$$

By the same logic, for any period n a shot is taken if $\tilde{p}_n > \text{cut}_{n+1}$. The cut point for period n is defined recursively as:

$$\text{cut}_n = \left(1 - \Phi\left(\frac{\text{cut}_{n+1} - \mu}{\sigma}\right)\right) * \text{E}(p|p > \text{cut}_{n+1}) + \Phi\left(\frac{\text{cut}_{n+1} - \mu}{\sigma}\right) * \text{cut}_{n+1} \tag{2.1}$$

since cut_N is pinned down by μ, we can recursively solve the system of equations defined by (1). The cut points (save the final one) are increasing in σ. This is because the expected value of the truncated normal distribution, for a given mean of the underlying normal distribution, increases with the variance. Intuitively since we are eliminating some of the low values in the support of the distribution, a higher variance increases the expected value because the probability of some high draws increases without concomitant increase in the probability in the corresponding low draws.

2.3.2 Examination of the assumptions

The model is designed to be a rough approximation to the actual shooting process. I do not claim that it fully captures the complexity of NBA shot selection, rather the intent is to provide testable predictions consistent with the theory. The following assumptions are maintained for estimation of the model, they are critically examined below:

[10] The shot allocation problem between players is addressed in Section 4.

1. Each possession has potentially N shooting opportunities

2. The shot opportunities arrive at intervals indexed by the shot clock

3. The team does not risk a turnover by continuing the possession

4. The shot opportunity distribution does not depend on the period

Failure of assumption (1) would make some low value shots appear to have been taken too early (i.e. "bad" shots) in the process. Consider the case where N=7 and there are 15 seconds left on the shot clock. In this case, the model holds that 4 opportunities remain, as such the cut point is fairly high. Suppose however that based on the "play" being run, if the shot is not taken only one opportunity truly remains because the offense will have to be "reset."[11] In this case the relevant cut is the N=7 threshold, far lower than the N=4. Many shots early in the process will look like bad decisions. As such, maintaining (1) will make it more difficult to confirm the model's predictions.

Assumption (2) is necessary for estimation as the shots need to be indexed by opportunities remaining. Of course using the shot clock is an imperfect way to order "potential" shots as it is uniform arrival is unlikely to hold. However, it probably approximately holds – more opportunities likely remain with 15 seconds as compared to 5 seconds on the clock. The underlying problem is that only opportunities that are acted upon are observed.

Assumption (3) is a consequence of assumption (1). Again it leads to some shots taken early in the process to appear sub-optimal. The reason being is that based on the defensive alignment the player might prefer to take a somewhat low-valued shot rather than risk a turnover.

Assumption (4) is probably violated mainly for shots taken within the first few seconds of the shot clock. In the data, roughly 60% of these shots immediately follow offensive rebounds, which usually means the distance is low and the defense is often out of position. Excluding these shots does not qualitatively affect the results of the section.

2.3.3 Observed data and estimation

The shot sample consists of 1,636 shots taken during the 2007/08 season. The expected point values were calculated as described in Section 2. Figure 1 shows a fractional polynomial curve fit of expected points as a function of the time remaining on the shot clock. The curve is overlaid with the density of shot times.

The key prediction of the optimal stopping model is that expected point value monotonically declines with time remaining on the shot clock. This prediction finds strong support in the data. As the time remaining on the shot clock decreases, the predicted point value of the shot does so as well. For instance, a shot taken with 15 seconds remaining is worth on average 1.36 points, whereas a shot taken with 10 seconds remaining nets only 1.29 points. This difference

[11] In basketball a play is a coordinated action by the offense designed to produce a good shooting opportunity.

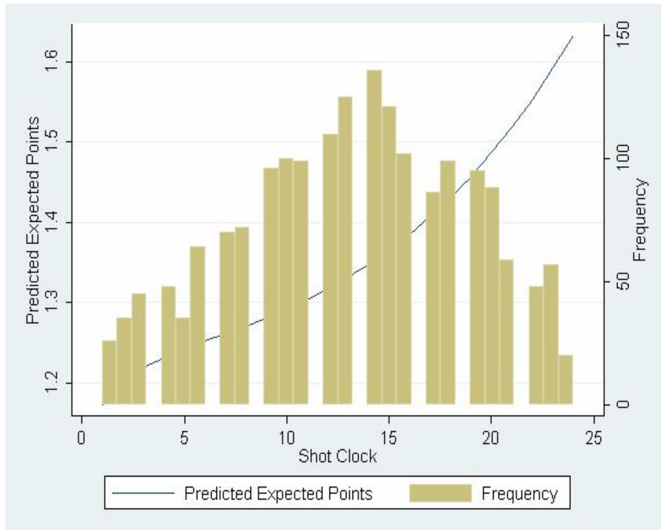


Figure 2.1: Observed shot selection and expected point values as a function of time remaining on the shot clock.

may not appear large at first but one must take into account that the average game has a little over 90 offensive possessions. As such a difference of .07 points per possession translates to a roughly 6.3 point difference at the end of the game - often the difference between winning and losing.[12] Of course this is not to say teams would be better off if players shot more quickly. The profile is declining precisely because players are willing to take lower valued shots later in the process due to the lower option value of continuing. Table 2 validates the statistical significance of the plot shown in Figure 1. Note that the result is robust to the exclusion of breakaways and put-backs (columns 3 and 4).

The model was fit to the data using nonlinear least squares (NLS). To do so, the data were placed in $N = 6$ shot clock bins, 1-4, 5-8, ... , 21-24. $N = 6$ was chosen because the fitted values implied a reasonable standard deviation, it provided fairly fine shot clock bins and the estimation led to realistic predictions in terms of expected points. The qualitative results of the estimation are not sensitive to the choice of N (for reasonable N, specifically the results are similar for N=4 through N=10).

Once the data was split based on shot clock, the optimal stopping curve was fit using non-linear least squares. The mean of the shot opportunity process is estimated to be 1.136 with a standard deviation of 0.336. The cut points and predicted frequency are given in Table 3:

Using the frequencies and cut points given in Table 3, the average points per possession predicted by the model is calculated to be 1.349, which is just below the observed average of 1.351.

[12]The Lakers average point differential for the season was +7.3 (source: espn.com).

Table 2.2: Expected Point Value as a Function of Time Remaining on the Shot Clock

	(1)	(2)	(3)	(4)
Sample	All		Omit breakaway/putbacks	
Time remaining	0.0171***	0.0977**	0.0160***	0.0667
	(0.0019)	(0.049)	(0.0022)	(0.061)
Time remaining^2		-0.0129*		-0.00674
		(0.0074)		(0.010)
Time remaining^3		0.000715*		0.000261
		(0.00043)		(0.00065)
Time remaining^4		-0.0000126		-0.00000170
		(0.0000086)		(0.000014)
Constant	1.132***	1.023***	1.143***	1.065***
	(0.028)	(0.10)	(0.029)	(0.11)
Excludes break-aways and putbacks	No	No	Yes	Yes
Observations	1636	1636	1527	1527
Wald Chi^2	79.04	88.92	54.93	64.95

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2.3: Cut Points and Frequencies, Baseline Case

Period	Cut	Predicted Freq
1	1.4711	0.181
2	1.4386	0.175
3	1.3982	0.170
4	1.3450	0.163
5	1.2686	0.155
6	1.1366	0.155

Intuitively, one could estimate the model by using the last period shot data to estimate μ and σ and then examining the implied cut profile. Doing so results in an NLS estimate of μ within .02 of the observed mean but a standard deviation estimate 0.1 less than the observed value 0.421. The reason for this that the relatively high standard deviation (0.421) would lead to higher cut points for the first few periods, which are not borne out in the data. This is probably due to a failure of assumption (1), some possessions offer fewer opportunities.

It is also worth noting that the optimal stopping rule predicts a fairly flat frequency of shots while the observed frequencies peak at about 15 seconds remaining on the shot clock. This peak is probably due to the running of a primary play (which takes some time to develop) and having fewer shot opportunities remaining than predicted by the model following the play.

The take home message from this section is that the prediction of point values monotonically decreasing with time remaining on the shot clock is found in the data. The shot clock might not always index shots in terms of remaining opportunities, but provided it gives a useful approximation the observed behavior is claimed to offer support for the optimal stopping model. Players appear willing to take lower valued shots later in the possession while shots taken earlier generally have higher value.

2.4 Shot allocation between players

In the previous section I abstracted from the problem of allocating shots between teammates, instead using a single team entity for simplicity. However, who takes a shot is just as important as when the shot is taken.

2.4.1 A simple model of shot allocation between players

In this section, the shot allocation problem between players is modeled as the offense choosing a vector of shooting probabilities (γ) for each of the five players. It is assumed that each player is endowed with a shot production function $\phi_i(\gamma_i)$ which is assumed to decrease in γ_i (it is also allowed to change with opponent). This assumption is designed to capture defensive adjustment via double-teams and traps. As a player shoots more, it becomes harder for him to find good shots. The continuity of $\phi(\cdot)$ is dependent upon the defense having a continuous adjustment mechanism. This is in fact the case provided the defense can randomize over adjustments such as double-teams, traps and zones. The fact that adjustments are binary variables (or more generally are finite) is overcome by randomizing over them continuously.

Payoff for the offense is given by:

$$U(\text{offense}) = \sum_{i=1}^{5} \gamma_i * \phi_i(\gamma_i) \tag{2.2}$$

Notably the offense's utility function has a risk neutral form - the team's objective is to maximize points per possession. I shall argue that this formulation is equivalent to maximizing the chance of winning the game. There is one obvious case where this does not hold. At the end of a close game the team that is behind may want to pass the ball to a player with a lower expected value but higher variance in his shot repertoire whereas the leading team might want to take a "safe bet." In both these cases the teams is making a mean for variance trade (albeit in different directions). To avoid this problem, the last two minutes of close games are eliminated from the sample if the score difference was less than 5.

Given the number of possessions per game (mean=91.7), lowering the average point value per possession in order to reduce or increase variance has a high probability of lowering the total points scored in the game. In a given match-up, call the team more likely to win the "favorite."[13] Holding mean constant, the favorite would like a lower variance of points per possession whereas the underdog team prefers higher variance. The intuition is that if they both perform as expected, the underdog will lose. As such the underdog has an incentive to increase the thickness of the tails of their point distribution whereas the favorite has precisely the opposite incentive. However, given the number of draws (possessions) the central limit theorem tells us that increasing per draw variance leads to about a 1/10th change in the variance of points for the entire sample (game). If a team sacrifices mean points per possession in order to increase/decrease variance it will lead to a lower point total with very high probability but a negligible change in variance.

In the sample the difference between the standard deviation in expected points of the most variable and consistent player was 0.16. Simulations contained in the Appendix 1.3 show that an underdog (favorite) would not want to lower point value per possession by even 0.01 points in exchange for a 0.16 increase (decrease) in standard deviation. Based on these observations, the risk neutrality assumption appears to be justified.

Provided more than one player shoots with positive probability, equilibrium in this game requires that each player (if they shoot at all) offers the same expected point value per shot. If any player on the offense had a higher point value on average the team would have this player shoot with increased frequency until the gains are erased – this reallocation from lower to higher point values increases utility (or equivalently, the probability of winning the game). Depending on the parameters, some players might never shoot because $\phi(0)$ is below the performance of other players in equilibrium. To have more than one player shoot with positive probability we simply assume there exists a j s.t. $\phi_j(0) > \phi_i(1)$ for all i.

2.4.2 Results

This model predicts that better players (those with higher ϕ_i profiles) shoot more but that all players net the same point values per shot. The implication is testable and the results

[13] Assuming this is common knowledge seems reasonable. Las Vegas sports lines for NBA basketball are very good at predicting who will win.

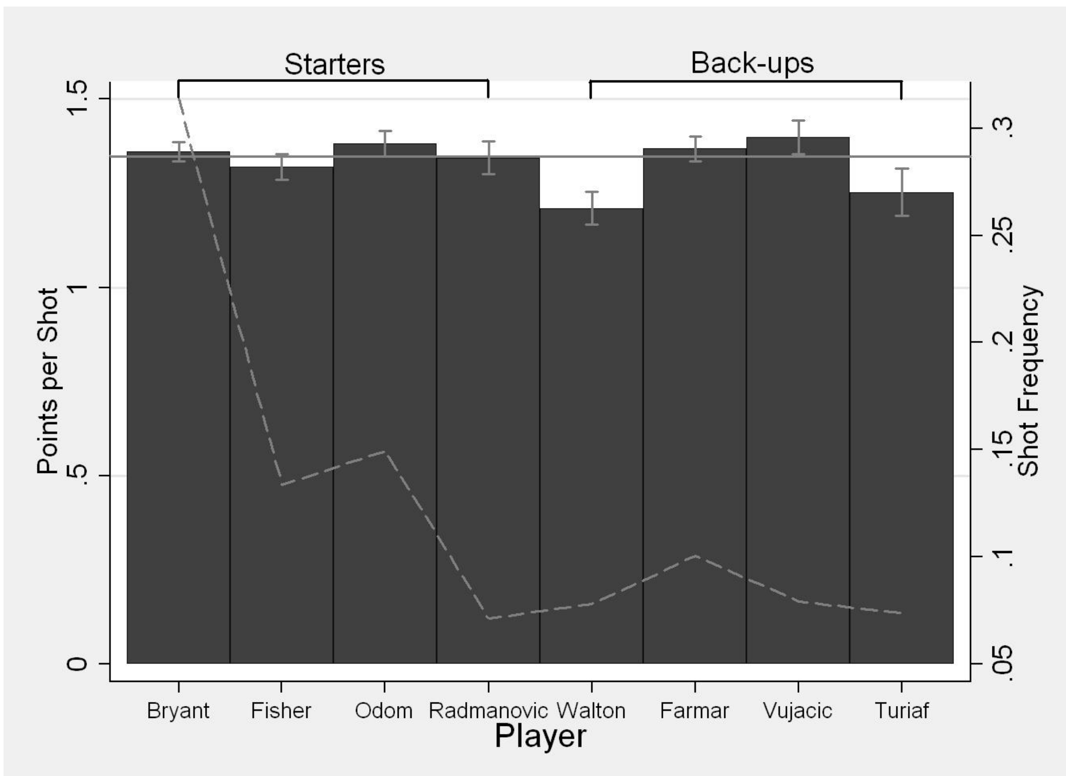


Figure 2.2: Expected point value by player.

are given in Figure 2. [14]

For the most part the expected point values are clustered around 1.375, all but 2 players are within .05 (approximately 2 standard errors) of this value. All 4 starters are within .02 – these players account for 67% of the teams total shots. There are 2 outliers, Turiaf and Walton, whose performance is substantially worse than their peers. Turiaf saw playing time during the season mainly due to injuries to two starters Andrew Bynum and Pau Gasol who were not included in the sample due to their injuries.[15] Furthermore, Turiaf was not offered a contract at the end of the season.[16] Walton is a back-up player whose playing time declined throughout the season, his performance indicates that the team should have allocated fewer shots to him. The outlier players account for only 14.5% of the teams total shots. Overall the prediction of equal shot values across players holds up well, especially amongst players in the starting line-up.

The right hand side axis of Figure 2 measures shot frequency. While shot value was constant across players, it is clear that some players bore more of the shooting load. The highest paid player (see appendix, Table A5), Kobe Bryant, took the highest percentage of the team's shots (31.3% across all shots). The second highest paid player, Lamar Odom, took the second highest percentage of shots. In an efficient labor market, the highest paid players will be the ones with the highest ability. Further evidence of these players' ability is that they have both been selected to the United States Olympic Basketball Team.[17] The model's prediction that high ability players shoot more, but with the same expected value as other players, finds strong support in the data.

The precise prediction of the model is that for any given five man line-up, each player offers the same expected point value. Figure 2 implicitly assumes that there are not significant line-up effects. We can relax this assumption and compare players when they are playing with a certain set of players. Given the number of line-up permutations, sample sizes will only be large for relatively common line-ups. Bryant, Fisher and Odom are the players who took the most shots, played the most minutes and were the highest paid. Table 4 limits the analysis to line-ups that included these "core 3" players.

Examining column (1), we see that the grouping of expected points is even tighter for the starting lineup with all players within 0.05 points of each other with values that do not significantly different from the midpoint of 1.365. For the case of the backup players, Vujacic and Turiaf fare slightly better when paired with the core three. The results indicate that Vujacic

[14] The reason there are only four starters is that many players started over the course of the season at the center position. The two players who were regular starters at the position, Andrew Bynum and Pau Gasol, did not amass enough shots to be part of the sample because Bynum had a season-ending injury shortly after earning the starting job and Gasol was acquired in a trade towards the end of the season.

[15] As sports writer Jeff Sibliksi writes "after Bynum went down with injury, Ronny [Turiaf] was forced to play as an undersized center until the team traded for Gasol. The shifting of positions limited Turiaf's game on both ends of the floor as he was often matched up against players that were much larger and taller than he."

[16] See http://fantasybasketball.usatoday.com/content/player.asp?sport=Nba&id=1167 for details

[17] Selection to the 12 man Olympic team has been considered a great honor since professionals were first allowed to compete in 1992, the year the famous "Dream Team" won the gold medal for the United States. Bryant is also an 11-time All-Star.

Table 2.4: Point values by player

Player	Mean	Obs	SE	Possessions per game
Starters				
Kobe Bryant	1.36	1027	0.015	59.18
Derek Fisher	1.33	530	0.019	50.76
Lamar Odom	1.38	535	0.019	58.89
Vladimir Radmanovic	1.36	238	0.026	44.04
Substitutes				
Jordan Farmar	1.36	223	0.024	20.94
Sasha Vujacic	1.44	176	0.031	23.20
Ronny Turiaf	1.21	203	0.039	34.07
Luke Walton	1.22	247	0.026	39.03

is slightly under-utilized when on the court with the core three starters although the difference is not significant at standard significance levels. In terms of adhering to the standards of optimality, we would expect the starting lineup to perform better than less frequently used alignments. This is indeed the finding of Table 5.

A regression framework can be employed to further examine the iso-expected point value prediction. Appendix Table A3 presents random effects and opponent fixed effects GLS regressions of expected points on dummy variables for each player, with Kobe Bryant as the omitted player. The coefficient estimates reproduce the findings presented in Figure 2 and Table 4. A comparison of the fixed and random effects estimators can provide additional insight. An implication of the theory is that the opponent should not affect the expected point value of shots by player. The reasoning is simple, if a player has a particular advantage against a certain team he should take a higher fraction of the teams shots until the gains are erased. This prediction holds up in the data as the coefficients for random and fixed effects are nearly identical. The associated Hausman-Wu test statistic is $1.02 \sim \chi^2_7$ with a p-value of 0.99.

Overall the data are quite supportive of the theoretical prediction of equal expected point values across players, especially within the starting line-up and when the analysis was restricted to lineups including the core three group of Bryant, Odom and Fisher.[18] The model presumed that performance was perfectly observable, in reality it might take a coach or teammates time to learn the level of a player's performance. It has already been mentioned that one outlier, Turiaf, saw playing time mainly due to injury and was the lowest paid player on the team. For the other outlier, Walton, we would expect his playing time to diminish as the season progressed due to his relatively low productivity. This is in fact the case. A simple linear regression of number of possessions played by Walton on a time (games ordered from first to last) yields a coefficient estimate of -0.204 ($t = 1.99$) indicating that over the course of the season his playing

[18]One might consider the findings of this section to be pure producer theory because it is basically a problem of allocation inputs (shots) to maximize output (points). However since the ϕ functions are motivated by the defensive response, I contend that there is also a game theoretic element at play.

time dropped by 12 possessions on average, or about 25%.

2.4.3 Changes in lineup

In the previous section I showed that while expected point value was relatively constant across players (especially starters) the "better" (those with higher salaries) players carried a larger percentage of the offensive load.[19] When the analysis was restricted to the core 3, the allocation decisions were in even closer adherence to the mixed strategy equilibrium prediction. However, we did see evidence that some players, especially backups, did in fact benefit from playing with a lineup which included the core three starters. It also might be the case that the inclusion of star players might raise the performance of players. In this case we would expect the increase to be uniform across players to maintain iso-expected value.

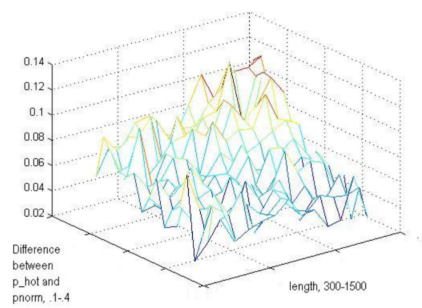


Figure 2.3: The effect of line-up composition.

The player most likely to positively affect teammates performance is 11-time All-Star and 2008 NBA Most Valuable Player, Kobe Bryant. In the context of the model, the positive effect is due to him taking the largest percentage of team shots, which attracts the defense away from his teammates leaving them with better shooting opportunities. Figure 3 is an analog to Figure 2 with the sample cut by Bryant's inclusion in the line-up. A clear pattern is evident. Most players show a uniform decrease in performance when Bryant is not on the court. The team still operates allocates the ball to achieve maximum productivity, but the overall level of productivity drops.

Appendix Table A4 confirms this assertion with regressions, by player, of expected point value on an indicator variable for Bryant's presence in the game. The analysis controls for opponents using fixed effects and supports the conclusion of Figure 3. The estimated coefficients are all positive and statistically significant for Odom and Farmar. Bryant's presence is estimated to have a sizable impact on the performance of all players except Radmanovic. The estimated impact is clustered around 0.07 points and as predicted by the model, is remarkably uniform

[19] Appendix Table A5 gives players' annual salary, contract length and contract expiry.

across players. The uniformity of the productivity decreases provides further support for the equilibrium predictions of the shot allocation model.[20]

2.5 Conclusion

This paper models shot selection and allocation in the game of basketball and examines the behavior of professionals relative to a developed theoretical standard of optimality. Analysis of the data reveals that the NBA players' decisions appear remarkably optimal. Shot selection over the course of a possession was modeled as an optimal stopping problem. This modeling decision was a natural one given that a shot must be taken before the 24 second shot clock expires. The model's prediction of monotonically declining expected point value with time remaining on the shot clock found strong support in the data.

A simple model of shot allocation across players, grounded in maximizing the probability of winning, generated the prediction that each player should have the same expected point value per shot, but that better players shoot with higher frequency. This familiar feature of mixed strategy Nash equilibrium found strong support in the data – for all but two players, shot value was uniform and not significantly different across players (within 0.025 points of star player Bryant). The results are strengthened considerably when limiting the analysis to the starting line-up. In these cases, which account for the majority of team shots, point value per shot was within 0.015 (0.75%) band across players. The two outlier players had significantly lower point value per attempt, but these back-up players took a relatively small fraction of the team's total shots. One of these players, Turiaf, was playing largely due to injury and his contract was not renewed, while the other, Luke Walton, experienced declining playing time throughout the season. A robustness check of the model showed that when star player Bryant was not on the court his teammates were uniformly less productive. That is, point value was still uniform but at a lower level.

The subjects of this paper are experts in a very high stakes industry. Their profession involves decision making in a fast-paced and complex game. Despite the underlying complexity of the game, their decision making closely adhered to the normative theoretical predictions. It is thus apparent that "NBA players got game theory."

2.6 Appendix

2.6.1 Further description of the data

The exact schedule of games used is available from the author. As are tapes and the corresponding spreadsheets for certain games that which were converted to digital format for

[20] Unreported regressions confirm that Bryant is the only player to (individually) have a statistically significant impact on the performance of this teammates. These regressions are available from the author.

those interested in studying the coding system. Due to copyright restrictions, I am unable to post the games online, however they can be distributed for private use. For the purposes of this study an offensive possession was defined as possession of the ball in the offensive half resulting in a shot. If the team had an offensive rebound this started a new possession. Buzzer beaters were omitted from the study (shots at the buzzer of length 30 ft or more).

2.6.2 Regressions

Table 2.5: Random-effects probit for shot success, by player

	K. Bryant	D. Fisher	L. Odom	J. Farmar	L. Walton	V. Rad	S. Vujacic	R. Turiaf
Distance	-0.199***	-0.245**	-0.351***	-0.372***	-0.233*	-0.116	-0.385*	-0.530***
	(0.062)	(0.12)	(0.10)	(0.12)	(0.13)	(0.17)	(0.22)	(0.16)
Distance^2	0.0139***	0.0198**	0.0255***	0.0280***	0.0104	0.00143	0.0239	0.0467***
	(0.0047)	(0.0091)	(0.0085)	(0.0098)	(0.011)	(0.014)	(0.017)	(0.017)
Distance^3	-0.0003***	-0.0005**	-0.0006***	-0.0006***	-0.0002	0.0000	-0.0005	-0.0013**
	(0.00010)	(0.00020)	(0.00020)	(0.00021)	(0.00027)	(0.00031)	(0.00036)	(0.00049)
Fouled	-0.806***	-0.529**	-1.004***	-0.407	-0.786**	-0.721*	-0.258	-0.346
	(0.11)	(0.25)	(0.17)	(0.30)	(0.32)	(0.37)	(0.37)	(0.23)
Turnaround	-0.00484	0.110	-0.207	-0.00931	0.509**	-0.528	0.269	-0.312
	(0.10)	(0.29)	(0.20)	(0.25)	(0.24)	(0.36)	(0.36)	(0.32)
Double team	-0.219*	-0.622*	-0.313	-0.227	-0.372	-0.562	-0.276	-0.0710
	(0.12)	(0.33)	(0.22)	(0.41)	(0.42)	(0.44)	(0.70)	(0.37)
Forced	-0.190	0.179	0.0649	0.0330	-0.193	0.221	-0.178	0.314
	(0.15)	(0.23)	(0.26)	(0.23)	(0.35)	(0.35)	(0.36)	(0.32)
Defended	-0.822***	-0.561***	-0.749***	-0.762***	-1.052***	-1.058***	-0.834***	-0.728***
	(0.13)	(0.17)	(0.21)	(0.23)	(0.28)	(0.38)	(0.26)	(0.28)
Defended	-0.336*	-0.708**	-0.600**	-0.399	-0.585*	-0.450	-1.253**	-0.744**
Layup	(0.19)	(0.33)	(0.24)	(0.35)	(0.35)	(0.45)	(0.62)	(0.35)
Defended	0.211	-0.370	-5.588	0.403	-5.225	0.556	-0.261	(dropped)
Three	(0.17)	(0.31)	(8652)	(0.36)	(8311)	(0.50)	(0.40)	(dropped)
Constant	1.503***	1.003***	1.746***	1.585***	1.624***	1.503***	1.962***	1.593***
	(0.20)	(0.37)	(0.23)	(0.33)	(0.33)	(0.40)	(0.66)	(0.31)
Observations	1404	597	665	449	348	316	357	330
Games	60	60	57	60	54	45	51	56

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The following table presents Probit estimates of the probability a shot is rebounded by the offensive team. While only one of the distance variables is statistically significant they are highly jointly significant with a χ^2_3 value of 16.43 (p-value=0.0009).

Table 2.6: Fixed Effects Probit Regression of Offensive Rebound

	(1)
	Offensive Rebound
distance	0.0859
	(0.055)
distance^2	-0.0137*
	(0.0076)
distance^3	0.000589
	(0.00038)
distance^4	-0.00000804
	(0.0000063)
defended	0.0379
	(0.076)
Double Team	0.0363
	(0.12)
Hard	-0.264***
	(0.090)
Constant	-0.439***
	(0.14)
Player fixed effects	Yes
Observations	2423[21]
Games	60

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.7: GLS estimates of expected point value on player

	(1)	(2)
Derek Fisher	-0.0414*	-0.0421*
	(0.022)	(0.022)
Lamar Odom	0.0217	0.0217
	(0.021)	(0.022)
Jordan Farmar	0.00766	0.00702
	(0.025)	(0.025)
Luke Walton	-0.150***	-0.152***
	(0.027)	(0.027)
Vladimir Radmanovic	-0.0156	-0.0159
	(0.028)	(0.029)
Sasha Vujacic	0.0366	0.0333
	(0.027)	(0.027)
Ronny Turiaf	-0.106***	-0.105***
	(0.028)	(0.028)
Constant	1.361***	1.361***
	(0.013)	(0.012)
Opponent fixed effects	No	Yes
Observations	4466	4466
Number of num_id	60	60

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2.8: The effect of Bryant's presence in the game on performance

	Fisher	Odom	Farmar	Walton	Radmanovic	Vujacic	Turiaf
Bryant=1	0.0675	0.106**	0.0762*	0.0623	0.0110	0.0690	0.0689
	(0.13)	(0.052)	(0.039)	(0.059)	(0.064)	(0.055)	(0.11)
Constant	1.254***	1.290***	1.324***	1.166***	1.336***	1.361***	1.200***
	(0.13)	(0.047)	(0.028)	(0.048)	(0.052)	(0.038)	(0.088)
Shots	597	665	449	348	317	357	329
Games	60	57	60	54	45	51	56

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.9: Los Angeles Lakers Salaries 2007/08 Season

Player	Salary	Contract Yrs Remaining	Yrs in League
Kobe Bryant	$19,490,000	3	11
Lamar Odom	$13,248,596	1	8
Vladimir Radmanovic	$5,632,200	1	6
Derek Fisher	$4,352,000	2	11
Luke Walton	$4,000,000	5	4
Sasha Vujacic	$1,756,951	3	3
Jordan Farmar	$1,009,560	2	1
Ronny Turiaf	$770,610	0	2

Source: hoopsworld.com, espn.com

2.6.3 Salaries

2.6.4 Risk Neutrality

The following plot shows the simulated winning percentage for an underdog with baseline mean expected point value of 1.38 playing a team who averages 1.4 points per possession with standard deviation 0.45. Each “game” was simulated 10,000 times. As evidenced by the figure, although the underdog wants to increase the standard deviation of the expected value of shot attempts it does not want to trade off any meaningful amount of mean to do so.

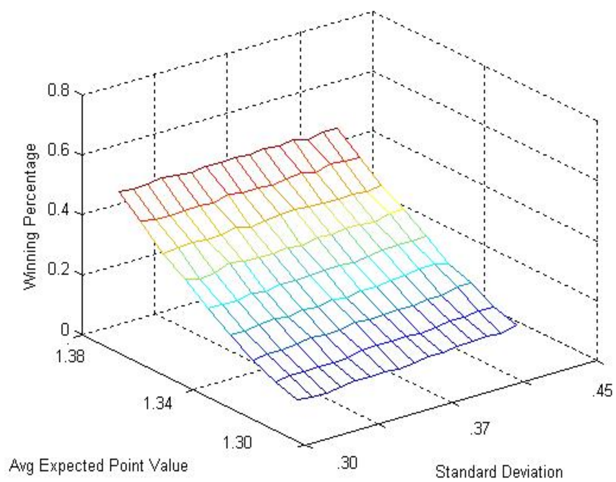


Figure 2.4: Underdog winning percentage as a function of standard deviation and mean

Chapter 3

Experts' Perceptions of Autocorrelation: The Hot Hand Fallacy Among Professional Basketball Players

Abstract

The hot hand fallacy is the perception positive autocorrelation when none is actually present. However, when mistaken beliefs do not affect the underlying success probability, then they come at zero cost if the underlying sequence truly exhibits independence. This paper studies an environment where success probabilities are a function of beliefs: shot selection in professional basketball. Beliefs are inferred through shot conditions such as time between shots, distance, defense, location, passes and touches. I find that a majority of the players in the sample significantly changed their behavior in response to hit-streaks by taking more difficult shots, while no player responded to miss-streaks. However, controlling for difficulty, shooting ability is not related to past outcomes. A quarter of the sample suffered a significant loss in shooting efficiency as a result of their mistaken responses. The remaining players either showed no response (3/8) or did so by substituting point value for difficulty leaving efficiency unchanged (3/8). Consistent performance is linked to heightened individual economic incentives.

3.1 Introduction

In many settings it is widely believed that past success is a good predictor of future success. Colloquially such sequential dependencies are referred to as having a "hot hand." There is a considerable evidence for belief in the hot hand. Basketball players and fans express the belief in surveys [Gilovitch et al., 1985, Burns, 2004], mutual funds often advertise they have beaten the market in past years[1] and patterns in sports betting are consistent with bettors believing a team "gets hot" [Camerer, 1989]. If there exists an underlying serial dependency in the data, belief in the hot hand perfectly rational. For instance, given the strong serial correlation in GDP, a macro economist would hardly quibble at the common expression "the economy is heating up." However, many studies have found that people tend to infer positive autocorrelation when none in fact exists. This inference mistake is known as the "hot hand fallacy" [Tversky and Kahneman, 1974].

However, studies which have found the hot hand fallacy have used novice experimental subjects or field subjects that had little to no financial incentive to have accurate beliefs. In contrast, this paper uses experts (professional basketball players) making decisions in a familiar environment (shot selection during games). In this setting, mistaken beliefs about serial dependence in ability can lead players to take shots that are too difficult, which can be costly for the team and individual. This paper addresses the questions: does the hot hand exist in professional basketball and do players respond appropriately to sequential dependencies (if any) in their ability? The goal is to further our understanding on the role of incentives in reducing inference mistakes and biases.

In a widely cited paper, Gilovitch, Vallone and Tversky (1985) study the hot hand in National Basketball Association (NBA) games. The authors present survey evidence that fans hold a strong belief in the hot hand.[2] The paper also surveys NBA players and finds that a majority (of the small sample) hold a belief in the hot hand. Contrary to the beliefs of players and fans, GVT fail to reject the hypothesis that shots are produced from a Bernoulli process with a constant success rate. In this paper I show that GVT's maintained assumption that players' shot selection is random (i.e. not dependent on past performance) is violated. If players take more difficult shots after a string of successful makes, then inference of autocorrelation in ability will be biased downwards if one fails to control for difficulty. A simple example is a player who is 5% better following a string of makes but adjusts by taking shots that are 5% more difficult – here one would erroneously find no hot hand effects.

Despite these shortcomings, the GVT finding has become a textbook example of a "behavioral bias." Agents, in this case basketball fans, (seemingly) commit mistaken inference and

[1] In fact such advertising drew the ire of the SEC and lead to a change in the rules governing permissible advertising practices. The SEC 2003-122 states: "The amendments address concerns that, especially in times of strong market performance, some funds may use advertising techniques focusing on past fund performance that may create unrealistic investor expectations or may mislead potential investors."

[2] Fans belief in the hot hand also been found by [Burns, 2004] which polled college students of varying basketball experience and [Rao, 009b] which examined the belief of "hardcore" fans found on Internet basketball message boards.

infer a pattern when none is in fact present. However, signal processing in a binary environment is a very difficult task and we might not expect unmotivated observers to do so following the edicts of Bayesian rationality.[3] On the other hand, professional players are the experts with multi-million dollar contracts. Inferring players' beliefs from questionnaires is unsatisfying from an economic perspective. Abstracting from the difficulty of eliciting true beliefs through a survey, economists are concerned with the behavioral implications of beliefs, not merely their subjective revelation. The standard method of analysis is to take observed behavior as the input and use it to define the set of action-consistent beliefs or preferences.

Using a rich dataset, I control for shot difficulty with shot conditions such as distance, defensive pressure and location and *find no evidence of the hot hand*. Given that shooting ability is not systematically related to past outcomes, I address the natural question: do players exhibit *through their actions* a belief in the hot hand, and if so, do the responses significantly impact shooting efficiency (points per shot)? Despite the fact that their ability does not predictably increase after made shots, some players shot selection indicates that they believe it does (i.e. they take significantly more difficult shots but perform no better than their baseline skill level). However such a response is not universal. A majority of the sample has consistent shot selection across past outcomes or responds in a manner that does not impact shooting efficiency.

Laboratory and field studies using relatively inexperienced subjects are very useful in that they allow us to indentify errors in reasoning endemic to human psychology. The shortcoming is that they are unable to examine if these biases are present in the face of strong countervailing incentives. Field studies using expert practioners allow us to examine how deep behavioral biases run. Professional sports offer an attractive testing ground for economic theory because the stakes are high and payoffs can be objectively quantified and observed on the individual level. Unlike most industries, here the productive activities of the firm are on full display.

The literature on inference mistakes in the perception of covariation is deep. Studies have found experimental subjects display a positive bias in the perception of covariance [Kareev, 1995], expect too much reversion in sequences (the "Gambler's Fallacy", Estes 1950) , lottery players are averse to choosing a number a week after it has won [Clotfelter and Cook, 1993, Terrell, 1994] and casino gamblers display both the hot hand fallacy and gambler's fallacy in roulette and craps [Croson and Sundali, 2005]. In all these cases, market forces do not provide a strong incentive for accurate inference. Experimental subjects are operating in an unfamiliar environment for low stakes, lottery players sacrifice a very small amount in expectation by avoiding previous winners[4] and all bets for given casino game have equal expected value. Further evidence of zero-cost mistaken perception of positive autocorrelation comes from financial markets. Sirri

[3] The weak power of classical statistical tests of binary outcome streaks, such as the runs test, speaks to this difficulty [Wardrop, 1995].

[4] In parimutuel lotteries a share of the proceeds from ticket sales is simply split among the winners, as such a player should try to choose numbers that others are averse to choosing. In non-parimutuel games there is no cost to mistaken beliefs. Players also make the mistake of not betting "unlucky" numbers such as 666 enough, and these mistakes are generally of a much larger magnitude than gambler's fallacy mistakes.

and Tufano (1998) find that investors are more likely to purchase top-performing mutual funds from the previous year despite the fact that there is little evidence of persistence in mutual fund performance.[5]

The dataset used in this paper comes from the 2007/08 Los Angeles Lakers NBA season.[6] The data set was constructed from NBA.com game logs and by actually watching the 60 games. It includes shot distance, defensive pressure, shot type, shot fouled, location, shot blocked, touches per player, time on the shot clock, double teams and if the shot was a turnaround or fadeaway. Observed shooting behavior following strings of makes and misses is analyzed to infer players' beliefs. I find heterogeneity of response to strings of successful shots; players fall into three groups: *sub-optimal responders*, *benign responders* and *consistent performers*. Both sub-optimal and benign responders react to past success by taking significantly more difficult shots, the difference is that benign responders do not show a significant drop in "points per shot" because they are shifting into 3 point shots, which offer increase value to offset the difficulty. Both groups of responding players also tend to pass the ball less and shoot more frequently. Consistent performers show no evidence of a belief in the hot hand.

The adjustments of the sub-optimal responders leads to significantly lower shot productivity and it is argued that these mistakes are of a costly, albeit not enormous, magnitude. That is they are large enough to alter game outcomes over the course of a season, but the net effect is likely in the range of one or two games (out of 82). I show that the players who sub-optimally respond to their perceived increase in ability had less economic incentive to optimally select shots due to contracts top-censored by the league's collective bargaining agreement. Whereas the players in the sample who adhere closest to the dictates of Bayesian rationality are young, unestablished players fighting for playing time and contracts.[7] The heterogeniety in response and link to economic incentives indicates that the while the hot hand fallacy is a robust behavioral bias, it is by no means universal or immune to incentives.

Of course to justifiably claim these predictable responses to past performance are mistakes, it must be adequately shown that ability following makes does not experience a concomitant increase.[8] Controlling for shot conditions, players show no evidence of ability changing as a func-

[5]Hendricks et al (1993) find evidence of persistence in mutual fund performance, but the persistence is generally short-lived and limited to the absolute best and worst funds in the sample. Carhart (1997) revisits the issue and concludes that overall the evidence does not support positive autocorrelation in mutual fund performance. Neither paper finds any evidence of negative autocorrelation. As such, the observed behavior of buying "hot" mutual funds comes at zero cost and potentially a small gain.

[6]The L.A. Lakers are one of teams in the National Basketball Association (NBA), which is the top professional basketball league in the world. NBA team salaries routinely exceed $80 million per team annually.

[7]NBA players enter the league on a standardized "rookie contract" the size of which depends on where they were selected in the annual draft. These contract are 3-4 years in duration. After this period, a player may become a "free agent" and sign contract dictated by his market value. Players who have performed well earn a contract often paying many times that of his rookie contract, while players who have not done so well might be unable to sign with any team and have to play in a lower paying league in Europe.

[8]I do not have to show that the hot hand does not exist per se, just that ability is not systematically related to past outcomes. There might be rare occurrences that players "get in the zone", but if they consistently respond to past outcomes by taking more difficult shots then ability should consistently respond as well. In section 3.3 I show that this is not in fact the case.

tion of past outcomes — this holds true for streaks of any length. I cannot reject that players in some rare instances have periods of higher ability, but it is shown that ability, in contrast to shooting behavior, does not systematically change as a function of past outcomes.

Interestingly, players respond far less to miss-streaks. Both groups of responding players show consistent performance across all streak lengths of misses. Their behavior after miss-steaks does not appear different than their baseline shot selection. For miss-streaks, all players in the sample have shot selection consistent with their observed performance as no player shows a decrease in ability following misses either. Inference mistakes in the serial dependence of ability are limited to the success domain. In the discussion section I present potential explanations for this asymmetry in response.

The rest of the paper proceeds as follows: Section 2 describes the data, Section 3 presents the main results of the paper, a discussion of the results follows in Section 4 and Section 5 concludes.

3.2 Data

The data used in the paper are 60 games from the Los Angeles Lakers 2007-08 National Basketball Association season. For each game, shot time, distance and outcome were first entered using the game logs (an online play-by-play including shot time and success/failure) available at espn.com and the NBA's official website.[9] Each game used in the study was recorded and the games were then watched by the author to record shot location, defensive pressure, touches for each player on the court, shot blocked/fouled, shot type and time on the shot clock.[10] The author performed all the game cataloging and did so in as objective a manner as possible.[11]

Table 1 describes the variables collected.

The Lakers were chosen because they have two players Kobe Bryant and Derek Fisher who are known around the league as streaky shooters. Bryant is considered by many to be the leagues top player and was the league MVP for the year of the study. Data was collected only for the Lakers and not their opponents. The NBA season is 82 games long, given the lengthy nature of the cataloging process (4-5 hours per game) it was determined that 60 games provided a large enough sample. The data set covers more games and is much richer than any previous study.

The appendix further describes the collection process.

[9] An example can be found at http://sports.espn.go.com/nba/playbyplay?gameId=280106013.

[10] For the first 34 games in the sample a binary variable was recorded when the shot clock was less than 5, for the remaining games the actual time was recorded.

[11] Many games have been transferred to .mpeg format and can be provided along with the coded spreadsheets to any parties interested in examining my cataloguing system.

[12] A shot is counted as defended if the defender is within 2ft of the shooter and actively guarding him when the shot is attempted, notably this excludes "close-outs" where a defender runs at an open shooter. The defended variable was the hardest to code, particularly difficult cases were watched many times. In these cases the motion of the shooter was examined in an effort to determine if the nearby defender affected the shot. Every effort was made to apply as objective a standard as the other measures used in the study.

Table 3.1: Descriptions of the variables

Variable	Description
Shot Time	Time of the FG attempt, includes attempts when the player was fouled
Defended[12]	=1 if player was guarded at the time of the shot
Hit	=1 if shot is made
Double team	=1 if player was defended by 2 or more opponents
Touches	Number of touches on a possession, by player
Team touches	Total touches on a possession
Shot Zone	One of 14 zones giving location on the court
Turnaround	Fadeaway, turn-around jumper or hook shot
Distance	Shot distance in feet, =0 for dunks and 1 for layups
Pulled	Player removed from game
Fouled	=1 if shooter was fouled
Blocked	=1 if shot is blocked
Forced	=1 if shot is taken with less than 5 seconds on the shot clock
Shot clock	time on the shot clock
Offensive rebound	=1 if the shot resulted in an offensive rebound

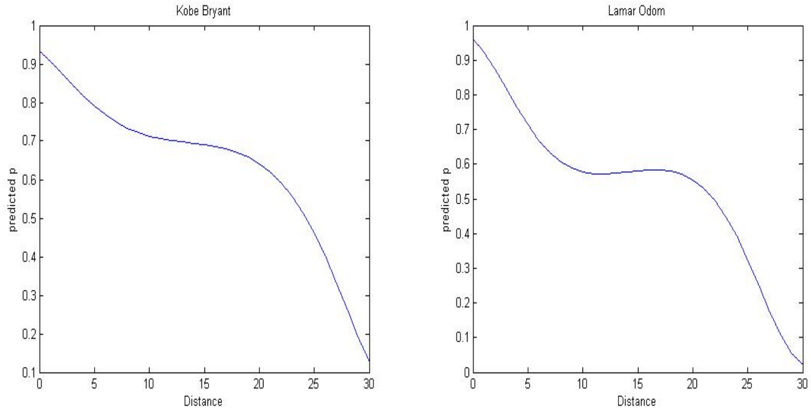
Table 3.2: Background on Players

Player	Description	Shots
Kobe Bryant	shooting guard, league MVP	1,418
Derek Fisher	point guard, good jump shooter	604
Lamar Odom	forward and low-post, decent jump shooter	677
Jordan Farmar	point guard, good jump shooter	455
Luke Walton	forward, “role player”	353
Vladimir Radmanovic	forward, long-range shooter	322
Ronny Turiaf	low-post, limited shooting range	334
Sasha Vujacic	shooting guard, Kobe’s backup	359

3.3 Results and analysis

The 8 players who took 300 or more shots during the sample period are included in the study. In the way of background, a basketball team consists of three types of players: guards, forwards and low-post. Guards bring the ball up the court and handle the ball on the perimeter during the offensive set. They are the primary ball handlers and long-range shooters. Low-post players are the tallest members of the team and generally take short-range shots near the basket and get rebounds. Forwards operate between the guards and the low-post. These distinctions are important because the range of shooting opportunities and accordingly the available responses to perceived “hotness” may differ across player types. Table 2 gives useful background information on the players in the study.

Figure 3.1: Background on Players



3.3.1 Predicting shot probabilities

For each player, the data has a panel structure by games (i) and possessions within games (t). As such, the natural estimating procedure to formulate predicted probabilities is random-effects Probit by player. Appendix Table 2 provides the estimation results. A cubic in distance is used to account non-linear increases in shot difficulty with distance. For some players none of the three distance terms are statistically significant but a joint F-test of significance overwhelmingly rejects the null of joint insignificance for all players. *Defended* is interacted with both *layup/dunk* and *three-pointer* to allow for differential effects of defense depending on shot type. Unreported regressions including the zone dummies were run but interestingly including shot location did not significantly improve fit for any player (controlling for a cubic in distance).

The coefficients have the expected signs. Notably being defended leads to a large and significant drop in shooting percentage. Most players had a lower shooting percentage when they took a "turnaround" type shot - since these shots were nearly always taken when being defended the marginal effects are not pronounced. Figure 1 shows the predicted success rates as a function of distance for undefended standard shots for two players: Kobe Bryant and Lamar Odom. Both patterns are typical, shooting percentage falls rapidly for the first few feet of distance, levels off for a mid-range "comfort zone" and then falls rapidly again for longer range shots. For low-post players this second fall occurs at about 12ft (2ft less than a free throw), while for long-range specialists it occurs as far as 25ft from the basket (1ft more than the longest three pointer).

3.3.2 Shooting performance as a function of past outcomes

The analysis of this subsection is done through Table 3 which relates shooting performance to past shot outcomes. The table is broken down by streak length and outcome (hit/miss). It allows for an easy comparison of behavior following a streak of X hits as compared to a streak of X misses. Note that the Table has telescoping structure. Columns 1 and 2 give all the outcomes after 1 hit and miss respectively, i.e. they include all shots taken by the player. Moving to columns 3 and 4 limits the sample to streaks of length 2 or more, and so on down the line.

A word of explanation is needed to understand the results presented in Table 3. "Hit rate" is the observed success rate for the shots in the column and "avg $\hat{p}$" is the average predicted success rate taken from the estimation of the previous subsection. Expected points for shot it are calculated as follows and then averaged across shots in the column:

$$\mathrm{E}_{it} = \begin{cases} \hat{p}_{it} * 2 + \hat{p}^{\text{off}}_{it}(1-\hat{p}_{it}) * \mathrm{EV}_{t+1} & \text{fouled=0, three=0} \\ \hat{p}_{it} * 3 + \hat{p}^{\text{off}}_{it}(1-\hat{p}_{it}) * \mathrm{EV}_{t+1} & \text{fouled=0, three=1} \\ \hat{p}_{it}(2 + f_i * 1) + (1-\hat{p}_{it}) * f_i * 2 & \text{fouled=1, three=0} \\ \hat{p}_{it}(3 + f_i * 1) + (1-\hat{p}_{it}) * f_i * 3 & \text{fouled=1, three=1} \end{cases}$$

where f_i is player i's free throw shooting success rate for the 2007/2008 season, EV_{t+1} is the expected value of a possession following an offensive respond, $\hat{p}$ is the predict shot success rate and $\hat{p}^{\text{off}}$ is the predicted probability of an offensive rebound for a given shot.

For the sample period the Lakers averaged 91.7 shots per game. Except in cases during the last minute of the game during close games, maximizing expected point value per shot is the optimal strategy because at the end of the game the team with the most points wins.[13] Rao (2009a) provides simulations that show that risk-neutrality maximizes the probability of winning the game. The intuition is that since there are a large number of possessions over the course of a game the law of large numbers implies that a risk-averse stance would lead to a much lower point total with probability near 1.

Significance levels shown in the tables are determined using t-tests which compare the hits section of each column with the miss section.[14] The observed success rate is never significantly different from the predicted performance.

The results are discussed player-by-player. The sheer amount of information can be initially overwhelming. Since Bryant takes over 1/3 of the shots when he is on the court and because he is the star player his table will receive the most attention.

[13] The last minute of play for games when the point margin was less than 5 were omitted

[14] There is one exception, the star on "42" in column 4 of Bryant's table. In this case the significance level indicates that Bryant has too few streaks of 4 makes given his overall shooting percentage. The significance was determined through simulations and maintains the assumption that success rate is constant. The result indicates that if we failed to account for shot difficulty we would (incorrectly) conclude that Bryant has a "oscillating" or "gambler's" hand.

Table 3.3: Shot selection by player as a function of past outcomes

Kobe Bryant								
	Consecutive Shots Made/Missed							
Outcome	1		2		3		4	
this period	Made	Missed	Made	Missed	Made	Missed	Made	Missed
Shots	645	706	291	387	119	205	42**	112
Hit Rate	0.422	0.424	0.399	0.434	0.353	0.459	0.405	0.438
Avg $\hat{p}$	0.408**	0.434**	0.399**	0.435**	0.382***	0.442***	0.410	0.422
Points	1.333**	1.394**	1.307**	1.397**	1.221***	1.379***	1.216	1.339
Derek Fisher								
	1		2		3		4	
Shots	240	304	91	157	31	78	11	52
Hit Rate	0.425	0.428	0.451	0.452	0.484	0.474	0.545	0.481
Avg $\hat{p}$	0.414	0.425	0.399	0.423	0.432	0.419	0.424	0.424
Points	1.291*	1.358*	1.190**	1.333**	1.170**	1.372**	1.195	1.443
Lamar Odom								
	1		2		3		4	
Shots	328	287	151	141	73	64	36	39
Hit Rate	0.451	0.443	0.477	0.468	0.521	0.438	0.556	0.436
Avg $\hat{p}$	0.444	0.465	0.446	0.481	0.430	0.463	0.454	0.509
Points	1.393	1.385	1.345	1.372	1.332	1.322	1.384	1.398
Luke Walton								
	1		2		3		4	
Shots	133	163	44	70	12	25	3	18
Hit Rate	0.368	0.472	0.364	0.529	0.333	0.600	0.333	0.556
Avg $\hat{p}$	0.394*	0.443*	0.394	0.444	0.367	0.424	0.263	0.438
Points	1.200	1.233	1.231	1.222	1.470	1.232	1.470	1.292

*** p<0.01, ** p<0.05, * p<0.1

Table 3.3 continued

Vladimir Radmanovic

	1		2		3		4	
Shots	130	146	49	68	26	29	13	18
Hit Rate	0.408	0.432	0.531	0.441	0.462	0.414	0.385	0.444
Avg $\hat{p}$	0.399**	0.444**	0.441	0.441	0.440	0.428	0.382	0.448
Points	1.324	1.375	1.379	1.357	1.339	1.343	1.301	1.359

Ronny Turiaf

	1		2		3		4	
Shots	125	149	44	74	15	38	4	27
Hit Rate	0.368	0.383	0.318	0.351	0.267	0.447	0.250	0.481
Avg $\hat{p}$	0.371	0.387	0.370	0.380	0.387	0.421	0.415	0.451
Points	1.238	1.324	1.174	1.322	1.341	1.335	1.332	1.302

Jordan Farmar

	1		2		3		4	
Shots	199	195	81	74	29	24	13	14
Hit Rate	0.427	0.503	0.383	0.514	0.517	0.375	0.615	0.429
Avg $\hat{p}$	0.463	0.477	0.448	0.493	0.494	0.496	0.496	0.493
Points	1.392	1.354	1.374	1.398	1.419	1.419	1.353	1.496

Sasha Vujacic

	1		2		3		4	
Shots	133	172	54	84	22	36	10	23
Hit Rate	0.451	0.436	0.463	0.452	0.409	0.472	0.200	0.478
Avg $\hat{p}$	0.419*	0.457*	0.388	0.440	0.395	0.479	0.380	0.471
Points	1.368	1.446	1.312	1.417	1.295*	1.504*	1.280	1.524

*** p<0.01, ** p<0.05, * p<0.1

In column 4 of Bryant's table, we see he has "too few" streaks of 4 hits. Using GVT's methodology of assuming random shot selection, we would conclude that Bryant has a statistically signficantly worse following makes (a "gambler's hand"). This result highlights the importance of controlling for shot difficulty. The reason he has too few streaks of 4 hits is that he takes progressively more difficult shots, based upon observables, after strings of hits. For instance after 3 hits his average predicted success rate is .382 while after 3 misses it is a healthy 0.442. Similarly after two hits average $\hat{p}$ is 0.399 while after two misses it is 0.435. These differences are highly significant. One might argue, however, that since the prediction model only took shooting conditions into account, and not past performance, that while the shots appear harder they are not in fact more difficult because ability has temporarily increased or decreased. If this were the case and the hot hand does in fact exist (at meaningful levels) we would expect actual performance to exceed predicted performance after hits and fall short of predicted performance after misses. For Bryant we note that this is not in fact the case —- the model performs very well. There are not systematic "model beating" tendencies that would point to the existence of a hot hand that predictably follows from past success. The following subsection addresses this point in much more detail.

The increased shot difficulty leads to a decline in points per shot for Bryant. In NBA competition, small differences in shooting percentage matter because games are often decided by only a few points. For example, in the 2007/08 season the difference between the best Western Conference team and the team that just missed the playoffs (8 teams make the playoffs per conference) was a mere 6.1 point differential per game and 18 (of 82) games played by the Lakers during the 2007/08 season were decided by 4 points or fewer.[15] In Table 7, Bryant's outside options are analyzed and it is shown that the net cost to the team should be adjusted down by approximately 0.05 points per shot because the defense increases pressure on all players during longer hit-streaks. As such the mistakes are not as costly as they initially appear, although they are still economically meaningful. In the discussion section I quantify the cost of these mistakes over the course of the season.

Unlike Bryant, Fisher does not take systematically more difficult shots in terms predicted success rate, but expected points do tend to decline after successes. The reason, which is shown in Table A1, is that Fisher draws fewer fouls after past success(es). This is because he is more likely to "settle" for a jump shot rather than driving the lane to draw a foul. Given his high free-throw shooting percentage, fouls provide high expected point value for the field goal attempt. Fisher and Bryant comprise the sub-optimal responder group.

The two remaining guards are Farmar and Vujacic both of whom respond less than their veteran counterparts. Farmar's expected shot value is stable across columns — he takes slightly harder shots, but these shots are 3-pointers offering higher returns to offset the difficulty (see Table 6). There also does not exist a pattern of exceeding or falling short of model predictions. Interestingly, Farmar is the least experienced player in the sample. An explanation is that since he is a young player fighting for playing time and hoping for a large free-agent contract, mistakes are far more costly. The lack of wiggle room creates a strong incentive for optimal behavior. Appendix Table 3 shows that there was wide range of annual salary and years remaining on contracts. Younger players yet to sign a large (by NBA standards) contract generally performed in closer adherence to Bayesian rationality than their veteran peers. I elaborate on this point in the discussion section.

Returning to the discussion of Table 3, Vujacic shows a pattern of taking more difficult shots, based upon observables, after makes than misses. For length 3 the model matches the observed hit rate for both cases and expected point value is significantly lower following 3 hits as compared to 3 misses. For length 4, the sample sizes are small — expected point value is lower following makes but the difference is not significant. His pattern of taking more difficult shots is not as pronounced as Bryant's as it only appears after strings of length 3 or more, whereas for Bryant even a single made shot has an effect.

In general forwards and low-post players respond less to past outcomes. One factor is

[15] Making the playoffs allows the team to host lucrative home games and boosts the player's reputation for future contracts. The higher ranked a team is during the regular season, the more playoff games it gets to host. This provides both direct financial windfall and the benefit of increasing the chances the team will proceed in the playoffs because of the home court advantage.

Table 3.4: Summary of Table 3

	Better than expected	Worse than expected	Sign Rank Test z-score
Following make(s)	19	13	0.90
Following miss(es)	17	15	0.90

that the unconditional variance in their success rate is lower as they have a more homogeneous shot repertoire. Another is that since they get far fewer touches per possession (see Tables 6 and A1), there are fewer opportunities to make a behavioral adjustment even if they held the hot hand belief.

Long-range shooter Radmanovic takes significantly more difficult shots following a hit as compared to a miss, but these changes do not lead to a significant decline in expected points because he is switching into 3-pointers (Table A1). His expected point values are stable across columns and the pattern of taking more difficult shots appears to only apply to the most recent shot. Walton has a similar pattern, he takes harder shots but experiences no drop in expected point value. Turiaf has stable shot difficulty and expected point values across columns. Odom displays a tendency to take slightly more difficult shots after hits as compared to misses, but in terms of expected shot value, his behavior does not trend up or down based up on past performance. These players make up the consistent performers group.

Interestingly, the patterns of response are driven only by streaks in the makes domain. For instance, the player who responded the most to his past success, Bryant, did not respond at all to miss-streaks. The predicted success rates of his shots for miss-streaks of length 1-4 were 0.44, 0.44, 0.43 and 0.45 respectively. His points per shot and other measures of shot selection given in Table 6 were stable as well. The results are similar for the remaining players in the sample.

3.3.3 Is ability systematically related to past outcomes?

One can observe from Table 3 that players do not exhibit a tendency to exceed the predictions of the model (based on observables alone) following makes as compared to misses. Put another way, there does not appear to be an unobserved change in ability. However, given the sheer amount of information contained in Table 3, it is useful to present aggregate analysis that makes the point more forcefully.

The sign rank score indicates that players beat the model by roughly the same magnitude following makes and misses. In absolute terms, 19 of the 32 observations (8 players by 4 lengths) exceed the model expectations following makes while 17 do so following misses. The reason the sign rank score is the same is that it is weighted by the magnitude of the differences, meaning that the differences for the case of makes were somewhat smaller than those of misses.

Another way to attack the question at hand is to put lagged performance into the

Table 3.5: Random-effects Probit Estimates of Shot Success on Lagged Performance

Player	Last Shot Made	Last 2 Made	Last 3 Made
Kobe Bryant	0.06	-0.01	-0.09
	(0.07)	(0.09)	(0.13)
Derek Fisher	-0.01	0.16	0.12
	(0.13)	(0.15)	(0.25)
Lamar Odom	0.03	0.09	0.34**
	(0.12)	(0.15)	(0.18)
Jordan Farmar	-0.22	-0.27	0.02
	(0.15)	(0.17)	(0.25)
Luke Walton	-0.17	-0.11	-0.05
	(0.17)	(0.23)	(0.46)
Vladimir Radmanovic	-0.08	0.36	0.07
	(0.19)	(0.23)	(0.31)
Sasha Vujacic	0.18	0.3	0.04
	(0.16)	(0.21)	(0.30)
Ronny Turiaf	0.03	-0.03	-0.24
	(0.18)	(0.26)	(0.43)
Total (+)	4	4	5
Total (-)	4	4	3
$\sum z - score$	-0.56	2.36	1.08
$\sum z^2$	5.42	8.76	2.87

*** p<0.01, ** p<0.05, * p<0.1

Probit model estimated in Appendix Table 2. Again, shot difficulty is controlled by the included conditioning variables. Table 5 reports the coefficient estimates and standard errors for "last shot made," "last 2 shots made," and "last 3 shots made" — which are added separately due to collinearity problems. If hot hand effects are systematically related to past outcomes, the estimates should display a positive trend (even if none are individually statistically significant). In fact such a trend is not present. The estimated coefficients are just as likely to be positive as they are negative and are generally quite close to 0.

The coefficient estimates are nearly evenly split between negative and positive. Out of 24 estimates, only 1 is significant at the 0.05 level, which is the expectation under chance. The second to last row gives the sum of the z-scores from the estimates. If there are significant underlying hot hand effects, then these values would be large and positive - they are not. The last row gives the sum of the squared z-scores, which under the null hypothesis that the coefficients equal to zero is distributed χ^2_8. The associated p-values are 0.81, 0.49 and 0.96 for columns (1), (2) and (3) respectively, which do not even approach conventional significance levels.

It is important to note that these results do not, and cannot, reject the notion that a player (very) occasionally drifts (i.e. unrelated to past performance) into a "hot state" and has period of improved performance. Rather the results show that ability is not *systematically* higher following makes as compared to misses. Conversely, it was shown in the past subsection that shooting decisions are, for a subset of players, systematically related to past outcomes and these changes lead to more difficult shots. We shall see in the following subsection that observable behavior is also systematically related to past outcomes and once again there is heterogeneity in the response ranging from none to quite large in magnitude.

3.3.4 Observable player behavior as a function of past outcomes

This section presents evidence on observed player behavior in the same format as the results from section 3.2. The purpose is two-fold. First, it provides a useful check of the results from 3.2 as it allows us to see the actual changes in shot conditions that led to the changes in difficulty and expected point value. Second, examining touches and the timing of shots allows us make further inference as to the beliefs of players.

Table 6 and Appendix Table 1 are the game conditions analog to Table 3. Given the sheer volume of information, the results are presented in the body of the paper for only two players (although they are discussed for all players), one sub-optimal responder and one benign responder. The remaining tables can be found in the appendix.

To understand the tables, recall that a "touch" is defined as holding the ball on a given possession in the offensive half of the court. For instance if a player brings the ball across half-court, passes and then later in the possession is passed the ball and shoots, then this counts as 2 touches for the player and 3 for the team. The results for touches include every possession during the streak, not just the ones the player took a shot. Hence sample sizes are much larger for this variable. A "one touch jumper" occurs when the player brings the ball up court and takes a shot that is not a lay-up or dunk (to rule out breakaways) without ever passing the ball. The decision of when to shoot over the course of possession is formalized in Rao (2009a) as an optimal stopping problem. The model predicts that shots are taken more quickly when the shooting opportunity offers a higher return. As such a player who believes he is hot may be more inclined to shoot the ball before passing to other teammates due to his perceived increase in ability. "Time gap" gives the elapsed game time in seconds between field goal attempts. Once again, statistical significance is determined using t-tests between makes and misses for a given streak length.

Table 3.6: Observable behavior as a function past outcomes

Kobe Bryant

	Consecutive Shots Made/Missed							
Outcome	1		2		3		4	
this period	Made	Missed	Made	Missed	Made	Missed	Made	Missed
One touch J	0.165**	0.126**	0.141	0.111	0.126	0.117	0.119	0.134
Touches	0.735***	0.679***	0.758***	0.677***	0.748**	0.660**	0.723*	0.624*
Distance	15.52***	11.76***	16.35***	11.98***	17.01***	11.73***	16.67**	12.19**
Turnaround	0.200	0.186	0.207*	0.158*	0.244**	0.151**	0.167	0.134
Three pointer	0.307***	0.166***	0.337***	0.178***	0.370***	0.156***	0.429***	0.179***
Defended	0.757	0.784	0.759	0.770	0.824	0.784	0.762	0.838
Fouled	0.146	0.177	0.134	0.171	0.101	0.161	0.048*	0.152*
Double team	0.089**	0.129**	0.076***	0.147***	0.084	0.127	0.071	0.134
Time gap	105.22**	120.7**	108.44	120.99	113.48	118.39	134.86	110.92
Obs	644	706	290	387	119	205	42	112

Jordan Farmar

	Consecutive Shots Made/Missed							
Outcome	1		2		3		4	
this period	Made	Missed	Made	Missed	Made	Missed	Made	Missed
One touch J	0.095	0.108	0.074	0.095	0.103	0.083	0.000	0.071
Touches	0.917	0.891	0.939	0.902	0.872	0.964	0.833	0.921
Distance	16.13	15.28	17.25*	14.07*	16.14	11.54	15.54	12.21
Turnaround	0.106	0.118	0.111	0.108	0.103	0.125	0.077	0.214
Three pointer	0.497	0.426	0.543**	0.378**	0.552	0.375	0.462	0.429
Defended	0.396	0.399	0.395	0.425	0.379	0.522	0.385	0.462
Fouled	0.075	0.046	0.074	0.068	0.034	0.083	0.000	0.143
Double team	0.020	0.026	0.012	0.014	0.000	0.042	0.000	0.000
Time gap	245.8*	312.94*	232.78**	398.41**	161.93*	298.29*	100.15*	278.5*
Obs	199	195	81	74	29	24	13	14

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Bryant is more likely to take a jump shot without passing after a single hit than a miss (significant at .05 level) but the difference is not significant for streaks of longer length. Bryant shots are roughly 4 ft longer following a hit than miss, this difference grows to over 6 ft for streak length 3. All the differences in distance are highly significant. Bryant is also significantly more likely to take a turnaround/fadeaway jumper when on hit streak as compared to a string of misses. Defensive attention does not seem to vary systematically across columns. However, Bryant does face significantly fewer double teams following hits as compared to misses. Initially this might seem counterintuitive, if the defense holds a hot hand belief they should double team more during hit-streaks. The finding is explained by the fact that most double teams occurs close to the basket and Bryant is taking longer shots on hit-streaks.

The touch data indicate that Bryant sees more of the ball during streaks of hits. His lowest touches per possession are during streaks of 4 misses, while they are highest after two makes, consistent with the notion that his teammates hold a belief in the hot hand. Time between shots is lower following 1 and 2 makes as compared to misses, but the difference is not significant for length 3 and in fact reverses sign for length 4. One possible explanation for the last finding is that defensive pressure steps up during long streaks (4 or more) forcing Bryant to wait longer between shots. Finally, Bryant is more likely to be fouled when on miss streak, the difference is significant at length 4, which can be attributed to taking closer range shots more likely to draw a foul. Overall the results for Bryant show that his behavior is significantly changing as a function of past outcomes. The differences are not only statistically significant but also of large magnitude.

Fisher, Vujacic and Farmar take more three-pointers following hits than misses. Fisher, like Bryant, is also more likely draw a foul when on a miss streak. Touches for these players do not seem to be systematically related to past outcomes. Farmar and Vujacic have lower time gaps when on a hit streak as compared to a miss streak.[16]

Odom's behavior with respect to shot conditions and touches is remarkably consistent across past outcomes. Radmanovic's is similarly consistent. Walton tends to take longer shots and significantly more three pointers, for example half his shots are three pointers after 3 makes, far in excess of his baseline propensity to take a three.

3.3.5 Outside options

The largest departures from optimal shot selection were by the best player, and league MVP, Kobe Bryant. One explanation for this has already been offered. A superior player has a lower private incentive to make optimal shot choices because he is not fighting for playing time or on cusp of the level of performance necessary to secure a lucrative "free agent" contract or

[16] This pattern of shooting naturally leads to temporally grouped hits. Due to memory constraints the temporal grouping can help explain hot hand beliefs among observers. Interested readers are referred to Rao (2009b) for a complete discussion.

Table 3.7: Expected shot values of teammates on the floor, by streak length

Streak length	Outcome	K. Bryant	D. Fisher
n/a	Off Court	1.288***	1.343
	On Court	1.346***	1.347
1	Missed	1.348	1.360
	Hit	1.348	1.336
2	Missed	1.347	1.366
	Hit	1.345	1.311
3	Missed	1.359	1.334
	Hit	1.323	1.293
4	Missed	1.370	1.329
	Hit	1.320	1.270

remain in the league. Salaries in second-tier leagues are far less lucrative than even the NBA minimum salary. In contrast, Bryant is signed to a censored "league maximum" contract.

There are two other possible explanations. The first is that Bryant does not hold a belief in the hot hand but his teammates do, so during a Bryant hit-streak they "lay back" and do not seek out good shooting opportunities for themselves forcing Bryant bear the offensive load. The second is that the defense increases overall pressure when he strings makes together because they either hold a hot hand belief, believe in team "momentum" or wish to avoid the Lakers getting too large a lead.[17] In either case, we would expect Bryant to take more difficult shots due to a decline in outside options (points per shot by his teammates on the court) and the increased personal pressure.

Table 7 summarizes the outside options for the 2 sub-optimal responders, Bryant and Fisher, during streaks of length 1-4. The first row simply compares the expected point value of teammates when the player is playing or not. The result indicates that Bryant does in fact "make his teammates better" as is commonly asserted.[18] This effect does not hold for Fisher. Rows 2 through 5 give the expected shot value of the other 4 players on the court broken down by streak length and hits/misses.

Table 7 shows that for streaks of length 1 and 2 outside options were very similar between makes and misses. For both players, the table does reveal a tendency for teammates' points per shot to be lower during strings of makes 3 or more. These differences reduce the net cost to the team — however, even after adjusting the magnitudes found in Table 3 down by the estimated margin (≈ 0.05 points), the statistical significance remains. It also lends support to the argument that defensive pressure applied to the whole team increases slightly when one player is on a streak of hits.

Comparing Tables 3 and 7 we notice that after 3 makes, Bryant's expected point value

[17] Note that if they increased pressure on just Bryant his teammates should have less pressure and therefore he should pass more and not experience a decline in performance. "Momentum" refers to a belief that a short run of good play by a team can have lasting effects in the game.

[18] This is effect is robust to controlling for the composition of players in the game.

was 1.22 for his own shots, while his teammates averaged 1.32 — the difference is statistically significant. After 3 misses the difference between Bryant and his teammates was less that .02. The pattern is similar for streaks of length 4. As such, even though there is a difference in teammate performance when Bryant is on a lengthy streak hit-streak, these differences are not large enough to explain Bryant's shot choices. The same analysis can be applied to Fisher, but no other players in the sample made these mistakes.

3.4 Discussion

In this section I discuss the heterogeneity in response across players, provide back-of-the-envelope calculations on the true cost to the team of sub-optimal shot selection by Bryant and Fisher and discuss the interesting result that players only respond to hit-streaks.

3.4.1 Heterogeneity in response and economic incentives

In the previous section it was shown that some players in the sample responded to past success by taking sub-optimally difficult shots. Explanations based on improved ability or increased defensive pressure on the whole team were ruled out. These players made mistakes that detracted from both personal and team performance. In contrast, some players exhibited remarkable consistency in their shooting behavior. For guards, the player with the least NBA experience (and smallest contract), Jordan Farmar, was the most consistent. The second lowest paid and experienced guard, Sasha Vujacic, also behaved more in line with the normative edicts of Bayesian rationality than his more experienced counterparts Bryant and Fisher.

There is an interesting link between the behavioral responses of the players and the economic incentives they faced. NBA contracts are subject to the terms of the Collective Bargaining Agreement (CBA) between the NBA Players' Association and the league owners. While the CBA is far too complex to describe in detail here, there are two key features that are important to the individual incentives faced by players in the sample.[19] The first is that contracts for players in their first 3 years in the league (such as Vujacic and Farmar) are set at pre-determined levels based on the player's selection number in the annual draft. These contracts are far less lucrative than those of even moderately successful veterans. A young player must perform well in order to sign a "free agent" contract, the terms of which are (generally) dictated by market value. Young players who do not perform at a high level are quite often unable to secure a contract at even the league minimum and are forced to take a large pay-cut in a second tier league. The second is that the CBA sets a maximum salary any player can earn.

Bryant earned the league maximum during the sample period, indicating that his true market value is probably above this level. In contrast, the young players of the same position

[19] Interested readers can find the entire CBA at http://www.nbpa.com/cba_articles.php.

had very high economic incentives. The breakdown of individual incentives due to the maximum contract can help explain Bryant's sub-optimal shot selection and the heightened incentives of the artificially low "rookie" contracts can help explain the fact that the young players did not experience a drop in shooting efficiency.

The link between incentives and behavioral biases has important implications. In many market settings, there are underlying behavioral biases that lead to a divergence of individual actions from firm interests. If biases are immutable, then enhancing incentives comes at a cost but generates no gain. However, if experienced agents can eliminate biases when properly incentivized, then firms should be mindful of underlying behavioral biases and structure contracts to punish agents who fall prey to them. The NBA maximum salary is similar to executive pay caps proposed in the wake of the 2008 financial crisis. For executives, the relevant bias is not the hot hand fallacy but rather present bias. It is in the interests of shareholders to structure contracts to eliminate present bias, but a pay cap places some executives in the position of an NBA player (such as Bryant) with a top-censored salary. High ability executives would face very liqttle economic incentive to overcome behavioral tendencies damaging to the firm. The results of this paper suggest that without proper incentives, latent biases are more likely to manifest themselves.

3.4.2 Quantifying costs

This subsection presents some conservative back-of-the-envelope calculations of the effective cost to the team of shot selection errors. In Table 3, I showed that two players, Bryant Fisher, took sub-optimally difficult shots following consecutive made shots. The most significant departures were streaks of length 3 and 4 (indeed these departures drove the differences in the whole table, given the telescoping structure of the presentation). For these streak lengths, Table 7 illustrated that the defense increased pressure on the entire team during these streaks. As such, the net cost to the team was lower, yet still significantly positive, than it initially appeared in Table 3.

For both players the net cost of their shot selection during hit-streaks of length 3 or 4 was approximately 0.12 points per shot. This value can used to calculate the aggregate cost. Bryant attempted 162 shots for streaks of length 3 or 4 and Fisher attempted 42. Using the value of 0.12 points per shot, this translates to about 25 total points for the 60 game sample or 34 point over the entire season. This estimate is conservative in that it excludes streaks of length longer than 4 (although there are very few of these). 34 points may not seem like a large value over the course of an 82 game season for a team that averages 90 points per game. However, NBA basketball is very competitive and many games are decided my slim margins. In the 2007/08 season the Lakers played 18 games that were decided by 4 points or less. As such, even using these conservative estimates it is likely that the shot selection mistakes cost the team 1 or 2 wins over the course of the season. For a playoff team, a single game usually determines if they get "home court advantage" — i.e. a single game translates to an economic loss through lost revenue in ticket sales.

3.4.3 The salience of the success domain

The results of Section 3.2, which showed that players tended to respond only to streaks in the makes domain (Section 3.4 showed some behavioral changes, but these did not impact shot difficult or shooting efficiency) . This result is consistent with the etymology and usage of the phrase "hot hand." Overestimating autocorrelation in general would also lead to a "cold hand" belief; however the focus of previous laboratory and field experiments and the colloquial usage is on past success predicting future success and not necessarily past failure predicting future failure. NBA players responses indicates that there is something more salient about successful outcomes that leads to mistaken inference.

The psychological salience of successful outcomes has intuitive appeal. It is a robust finding that people tend to accentuate the positive and downplay the negative in their recall of events and attributional judgment. There is also evidence that of hardwired optimism in our updating machinery. Eil and Rao (2009) show that subjects update there beliefs differently when the signal they receive is good as opposed to bad news. The authors use an objective signaling environment in which subjects receive binary signals (up/down comparisons to another subject) about there intelligence or physical attractiveness. They find that subjects tend to respond more to good news than bad news. This result can help explain why (some) players respond to hit-streaks of length 3 and 4 but not to miss-streaks of the same length.

3.5 Conclusion

This paper uses a rich data set to examine the beliefs of NBA players concerning the hot hand. The underlying methodology is that beliefs should be inferred from observed behavior. Using shot difficulty, touches per possession, passing behavior, and time between shots I find heterogeneity in players' hot hand beliefs. Broadly the players fall into three categories: *sub-optimal responders*, *benign responders* and *consistent performers*. The players who do respond to their past performance tended to take longer shots, were less likely to draw a foul, touched the ball more and passed the ball less following a make or string of makes. Interestingly, there is no evidence that any player responded to strings of misses. The analysis of Section 3.3 showed that, controlling for shot difficulty, players did not exhibit an increase in ability following strings of made shots or a drop in ability following misses; i.e. there is no serial dependence in ability (the hot hand does not exist for the players in my sample).

I argue that the significant change in behavior provides evidence of a belief in the hot hand. Given that players do not experience a concomitant increase in ability, the taking of difficult shots following a string of makes lead to a significant drop in expected shot value for *sub-optimal responders* (2 of 8 players). The most notable player in this group is 2007/08 League MVP Kobe Bryant. In sharp contrast, *consistent performers* (3 of 8 players) did not systematically respond to hit-streaks and as such do not suffer from reduced performance. These players observable actions are remarkably stable across past outcomes. A third group, *benign responders* (3 of 8), did significantly respond to past success by taking more difficult shots — but did so by switching into 3-pointers that provided increased returns, which offset the increased difficulty leaving points per shot stable. There is evidence that players with higher economic incentives (i.e. those with a low margin of error in performance) were more likely to occupy the latter two groups.

The finding that some players do in fact commit costly shot selection mistakes sets the result apart from other evidence of behavioral biases in Bayesian updating. These players exhibit a bias previously found amongst basketball fans, experimental subjects and field subjects (lottery, casino) but in contrast to these previous studies, here exhibiting the bias is costly. The result shows the robustness of the hot hand fallacy in human psychology. However, the fact that the players with greater economic incentive to perform consistently did not exhibit the bias indicates that it is by no means universal. Generalizing the results to other settings, they suggest that contracts should be structured to help eliminate behavioral biases.

3.6 Appendix

3.6.1 Further description of the data

The exact schedule of games used is available from the author. As are tapes and the corresponding spreadsheets for certain games saved on the author's hard drive. For the purposes of this study an offensive possession was defined as possession of the ball in the offensive half resulting in a shot. If the team had an offensive rebound this started a new possession. If readers thought the touch numbers were lower than they might expect this is why. For instance, on a put-back the possession would only have one touch.

Buzzer beaters were omitted from the study (shots at the buzzer of length 30 ft or more).

Table 3.8: Table 3 extended

Derek Fisher

Outcome this period	Consecutive Shots Made/Missed 1 Made	1 Missed	2 Made	2 Missed	3 Made	3 Missed	4 Made	4 Missed
One touch J	0.154	0.191	0.187	0.197	0.323	0.231	0.182	0.250
Touches	0.810	0.800	0.790	0.824	0.734	0.804	0.811	0.757
Distance	17.7***	15.72***	17.22	16.62	17.1	16.87	17.73	16.77
Turnaround	0.054	0.036	0.067	0.038	0.100	0.038	0.000	0.019
3 pointer	0.427***	0.309***	0.367	0.363	0.267	0.346	0.364	0.327
Defended	0.456	0.452	0.522	0.465	0.484	0.474	0.455	0.462
Fouled	0.059*	0.102*	0.022*	0.083*	0.000*	0.115*	0.000	0.154
Double team	0.046	0.043	0.056	0.032	0.033	0.026	0.091	0.019
Time gap	230.28	225.2	257.38	228.27	323.13	270.24	436.18	285.56
Obs	239	304	90	157	30	78	11	52

Lamar Odom

Outcome this period	Consecutive Shots Made/Missed 1 Made	1 Missed	2 Made	2 Missed	3 Made	3 Missed	4 Made	4 Missed
One touch J	0.018	0.017	0.020	0.014	0.027	0.016	0.000	0.000
Touches	0.534	0.567	0.522	0.557	0.540	0.524	0.560	0.510
Distance	7.8	7.58	9.03	7.37	9.63	8	9.81**	5.67**
Turnaround	0.101	0.091	0.099	0.128	0.110	0.125	0.139	0.103
3 pointer	0.137	0.115	0.159	0.113	0.151	0.156	0.139	0.051
Defended	0.657	0.644	0.622	0.626	0.634	0.656	0.571	0.667
Time gap	241.16	209.65	252.27	191.74	228.81**	182.89**	211.5	164.49
Fouled	0.180	0.153	0.139	0.128	0.151	0.109	0.167	0.128
Double team	0.098	0.066	0.099	0.064	0.068	0.047	0.056	0.000
Obs	328	287	151	141	73	64	36	39

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.8 (cont.) Table 3 extended

Luke Walton

Outcome	Consecutive Shots Made/Missed 1		2		3		4	
this period	Made	Missed	Made	Missed	Made	Missed	Made	Missed
One touch J	0.008	0.025	0.023	0.014	0.000	0.000	0.000	0.000
Touches	0.539	0.553	0.502	0.510	0.474	0.516	0.609	0.520
Distance	11.32**	9.21**	13.30***	8.41***	14.17	9.84	17.33	10.33
Turnaround	0.135	0.123	0.159	0.100	0.083	0.000	0.000	0.000
3 pointer	0.195	0.160	0.250*	0.129*	0.500**	0.160**	0.667*	0.167*
Defended	0.515	0.547	0.409	0.580	0.333*	0.520*	0.333	0.444
Fouled	0.113	0.067	0.114	0.057	0.250	0.080	0.333	0.111
Double team	0.030	0.043	0.068	0.071	0.250	0.080	0.333	0.056
Time gap	379.58**	241.11**	366.05	261.33	206.75	197.08	132.33	226.94
Obs	133	163	44	70	12	25	3	18

Vladimir Radmanovic

Outcome	Consecutive Shots Made/Missed 1		2		3		4	
this period	Made	Missed	Made	Missed	Made	Missed	Made	Missed
One touch J	0.031	0.048	0.041	0.044	0.038	0.069	0.077	0.056
Touches	0.434	0.441	0.486	0.421	0.545	0.446	0.607	0.461
Distance	15.67	14.21	14.98	14.09	13.81	16.03	13.23	14.72
Turnaround	0.085	0.068	0.041	0.088	0.038	0.103	0.077	0.167
3 pointer	0.523	0.479	0.531	0.515	0.500	0.621	0.462	0.556
Defended	0.430	0.427	0.375	0.455	0.440	0.393	0.500	0.412
Fouled	0.092	0.068	0.082	0.044	0.077	0.034	0.154	0.056
Double team	0.046	0.048	0.020	0.059	0.038	0.103	0.077	0.056
Time gap	289.61	271.25	255.92	264.43	207.81	268.55	192.69	341.44
Obs	130	146	49	68	26	29	13	18

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.8 (cont.): Table 3 extended

Ronny Turiaf

Consecutive Shots Made/Missed

Outcome	1		2		3		4	
this period	Made	Missed	Made	Missed	Made	Missed	Made	Missed
One touch J	0.008	0.013	0.000	0.027	0.000	0.026	0.000	0.037
Touches	0.428	0.394	0.403	0.393	0.361	0.382	0.286	0.370
Distance	8.71	5.92	8.75	5.11	6.53	4.89	4.75	5.3
Turnaround	0.080	0.081	0.045	0.135	0.133	0.105	0.250	0.037
3 pointer	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Defended	0.610	0.703	0.674	0.743	0.733	0.684	0.750	0.63
Fouled	0.368	0.242	0.182	0.243	0.267	0.211	0.250	0.148
Double team	0.032*	0.081*	0.023	0.068	0.000	0.053	0.000	0.074
Time gap	327.26	308.6	378.82	287.93	266.33	313.37	116.5	350.78
Obs	125	149	44	74	15	38	4	27

Sasha Vujacic

Consecutive Shots Made/Missed

Outcome	1		2		3		4	
this period	Made	Missed	Made	Missed	Made	Missed	Made	Missed
One touch J	0.045	0.029	0.037	0.024	0.000	0.028	0.000	0.043
Touches	0.564	0.573	0.557	0.599	0.613	0.517	0.630	0.531
Distance	20.05	19.15	21.63**	18.06**	20.77	17.28	20.5	17.96
Turnaround	0.075*	0.029*	0.074	0.036	0.091	0.028	0.200	0.043
3 pointer	0.586	0.570	0.685*	0.524*	0.591	0.556	0.500	0.565
Defended	0.364	0.281	0.389	0.333	0.409	0.250	0.500	0.261
Fouled	0.053	0.052	0.037	0.071	0.045	0.056	0.100	0.087
Double team	0.008	0.023	0.019	0.036	0.000	0.056	0.000	0.043
Time gap	205.06**	303.92**	206.26	272.79	257	290.22	91.1	315.04
Obs	133	172	54	84	22	36	10	23

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.9: Random-effects probit for shot success, by player

	K. Bryant	D. Fisher	L. Odom	J. Farmar	L. Walton	V. Rad	S. Vujacic	R. Turiaf
Distance	-0.199***	-0.245**	-0.351***	-0.372***	-0.233*	-0.116	-0.385*	-0.530***
	(0.062)	(0.12)	(0.10)	(0.12)	(0.13)	(0.17)	(0.22)	(0.16)
Distance^2	0.0139***	0.0198**	0.0255***	0.0280***	0.0104	0.00143	0.0239	0.0467***
	(0.0047)	(0.0091)	(0.0085)	(0.0098)	(0.011)	(0.014)	(0.017)	(0.017)
Distance^3	-0.0003***	-0.0005**	-0.0006***	-0.0006***	-0.0002	0.0000	-0.0005	-0.0013**
	(0.00010)	(0.00020)	(0.00020)	(0.00021)	(0.00027)	(0.00031)	(0.00036)	(0.00049)
Fouled	-0.806***	-0.529**	-1.004***	-0.407	-0.786**	-0.721*	-0.258	-0.346
	(0.11)	(0.25)	(0.17)	(0.30)	(0.32)	(0.37)	(0.37)	(0.23)
Turnaround	-0.00484	0.110	-0.207	-0.00931	0.509**	-0.528	0.269	-0.312
	(0.10)	(0.29)	(0.20)	(0.25)	(0.24)	(0.36)	(0.36)	(0.32)
Double team	-0.219*	-0.622*	-0.313	-0.227	-0.372	-0.562	-0.276	-0.0710
	(0.12)	(0.33)	(0.22)	(0.41)	(0.42)	(0.44)	(0.70)	(0.37)
Forced	-0.190	0.179	0.0649	0.0330	-0.193	0.221	-0.178	0.314
	(0.15)	(0.23)	(0.26)	(0.23)	(0.35)	(0.35)	(0.36)	(0.32)
Defended	-0.822***	-0.561***	-0.749***	-0.762***	-1.052***	-1.058***	-0.834***	-0.728***
	(0.13)	(0.17)	(0.21)	(0.23)	(0.28)	(0.38)	(0.26)	(0.28)
Defended	-0.336*	-0.708**	-0.600**	-0.399	-0.585*	-0.450	-1.253**	-0.744**
Layup	(0.19)	(0.33)	(0.24)	(0.35)	(0.35)	(0.45)	(0.62)	(0.35)
Defended	0.211	-0.370	-5.588	0.403	-5.225	0.556	-0.261	(dropped)
Three	(0.17)	(0.31)	(8652)	(0.36)	(8311)	(0.50)	(0.40)	(dropped)
Constant	1.503***	1.003***	1.746***	1.585***	1.624***	1.503***	1.962***	1.593***
	(0.20)	(0.37)	(0.23)	(0.33)	(0.33)	(0.40)	(0.66)	(0.31)
Observations	1404	597	665	449	348	316	357	330
Games	60	60	57	60	54	45	51	56

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.10: Los Angeles Lakers Salaries 2007/08 Season

Player	Annual Salary	Contract Yrs Remaining	Yrs in NBA
Kobe Bryant	$19,490,000*	3	11
Lamar Odom	$13,248,596	1	8
Vladimir Radmanovic	$5,632,200	1	6
Derek Fisher	$4,352,000	2	11
Luke Walton	$4,000,000	5	4
Sasha Vujacic	$1,756,951	3	3
Jordan Farmar	$1,009,560	2	1
Ronny Turiaf	$770,610	0	2

* Maximum Permitable. Source: hoopsworld.com, espn.com

Chapter 4

When the Gambler's Fallacy becomes the Hot Hand Fallacy: An Experiment with Experts and Novices

Abstract

The "gambler's fallacy" (GF) and "hot hand fallacy" (HHF) are two related inference mistakes – GF is a belief in too little serial correlation, while HHF is a belief in too much. Theoretical models have shown that both can be explained by an underlying belief in the "law of small numbers." This study uses a subjective probability sequence (professional basketball shooting) to see if GF and HHF can both exist within a single agent. Subjects were experienced basketball fans or novices. Across subjects, there is statistically significant GF after short streaks and HHF after long streaks. On the individual level, a transition from GF to HHF is most common. Consistent GF or HHF was rare and the expert sample generally exhibited bias of a lower magnitude as compared to the novice sample.

4.1 Introduction

The gambler's fallacy (GF) is a belief that a streak (consecutive occurrences of an outcome) of a given length is more likely to end than dictated by Bayesian inference, i.e. a underestimation of the degree of autocorrelation. The hot hand fallacy (HHF) is the inverse belief – a streak is believed to continue with irrationally high probability. There is considerable evidence for both these inference mistakes, yet they appear to be in direct opposition to each other. In an effort resolve this apparent inconsistency, researchers in both psychology and economics have argued that both biases are grounded in a belief in the "law of small numbers" [Tversky and Kahneman, 1974, Rabin and Schrag, 1999].

A common root cause generates an interesting testable implication: if an agent believes in the law of small numbers then he should show evidence of both GF and HHF. The biases are not in competition, rather they complement each other and which one will be operational is situationally dependent. I test this hypothesis using a subjective probability sequence: shots in a professional basketball game. Beliefs were inferred from the subjects' predictive bets. The majority of subjects exhibited GF for certain past realizations and HHF for other sets of past outcomes – the transition between the two biases was a predictable function of streak length. The results show that GF and HHF can coexist and add credence to the argument that they are based in the same underlying reasoning mistake. A consequence is that it is not very informative to identify a person as GF-type or HHF-type. Rather the importance should be placed on the cognitive process, which is argued (following other authors) to be a use of the "representative heuristic," and the environmental factors that determine the operational bias.

GF was first identified in the laboratory in the "probability matching" experiments of the 1950's, although the etymology indicates the notion predates laboratory study [Estes, 1950]. Field evidence of GF includes betting in parimutuel lotteries [Terrell, 1994, Clotfelter and Cook, 1993] and roulette wagers in casinos [Croson and Sundali, 2005]. There is also strong evidence in the literature for the HHF. Mutual funds have been shown to have independent returns from year to year [Carhart, 1997] yet investors have a preference for "hot" funds [Sirri and Tufano, 1998]. Pro basketball shooting skill has been shown to be serially uncorrelated [Gilovitch et al., 1985, Rao, 009aa] but National Basketball Association (NBA) players take sub-optimally difficult shots following a string of makes [Rao, 009aa] and spectators show a strong belief that skill level is affected by past shot outcomes [Gilovitch et al., 1985].

Tversky and Kahneman (1971) first argued that both GF and HHF could be explained by the representativeness heuristic – the belief that small samples should match their parent population (it is also known as the "law of small numbers") . Rabin (2002) and Rabin and Vayanos (2009) formalize this argument mathematically. The intuition behind the Rabin model is that belief in "law of small numbers" generates a belief in frequent outcome alternation (GF). This is because consecutive occurrences of a given outcome (streaks) are unrepresentative of the parent population so they are expected to end. However, long streaks violate this expectation of alternation, to reconcile the observed outcomes the success rate of the underlying Data Generating Process (DGP) is revised. This pattern of revision leads to a belief in positive autocorrelation because strings of successes lead to upward revisions in the success rate, which helps explain the "surprising" (to the agent) lack of alternation observed in the sequence.

The model predicts that there should be a transition from GF to HHF when there is uncertainty about the underlying DGP. Sequences generated with a subjective probability measure (e.g. mutual fund performance, shooting accuracy) are better candidates for model revision than those generated from objective probabilities (e.g. spins of a roulette wheel). As such, we may not observe the GF to HHF transition in the objective cases. Consistent with this reasoning, Croson and Sundali (2005) find that after streaks of 5 or more roulette outcomes (colors) casino gambler's showed significant GF (for streaks of 4 or less no biases were evident) but not HHF. More generally the field evidence which for HHF typically involves subjective processes such as mutual fund returns while evidence supporting GF involves objectively random processes such as draws from urns in a laboratory. This pattern also finds support in the laboratory [Burns and Corpus, 2004].

In this experiment, subjects made outcome predictions of shots taken in a NBA game. The game tape was edited so that a "freeze-frame" displayed just as the ball left the shooter's hand. During this interval, subjects placed a bet on the outcome of the shot, with payoffs weighted by the shot's objective difficulty in a method described in detail in section 2. The payoffs were designed to mimic returns to stocks on an efficient exchange – subjects were incentivized to bet "hit" when they felt the shooter was more likely than usual to make that particular shot (i.e. conditional on distance, defense, etc.) and bet "miss" if they felt the shooter was less likely to make the shot than their performance on that type of shot would dictate.

Basketball shooting is a sequence traditionally thought to lead to the HHF (Gilovitch et al. 1985), yet I find significant GF for short streaks and HHF for longer streaks, especially among the subjects who reported that they did not watch or play basketball (novice group). Subjects who reported watching more than 30 basketball games per season (expert group) exhibited the same patterns but with lower magnitude and lacking statistical significance. However both groups display a statistically significant relationship between the two biases. While most subjects displayed GF for short streaks and HHF for long streaks, there was a significant minority with precisely the opposite pattern. Indeed subjects who consistently displayed either GF or HHF across streak lengths were rare. The result provides strong support for the common root cause hypothesis, shows the interdependence between the two sets of beliefs and highlights the importance of context in discussing behavioral biases.

4.2 Experimental Design

The experiment was conducted at the University of California San Diego economics laboratory. Upon entry to the subject database, potential participants were asked background questions about the amount of basketball they typically watched. Two groups of subjects were invited to participate in the study: those who reported watching more than 30 games per season and those that reported watching 0.[1] These groups constituted the "expert" and "novice" sub-

[1] A post-questionnaire confirmed the answers to these questions and was used to place subjects in each group.

samples respectively. 70 subjects (39 experts and 31 novices) participated across six 1-hour sessions, earning $14.33 on average.

Subjects viewed game tape from the 2007/08 Los Angeles Laker's season. The tape was edited so that for some shots, referred to as "payoff shots," a freeze-frame was displayed for 10 seconds just as the ball was leaving the shooter's hand. During this interval subjects bet either "hit" or "miss" and chose to wager either 1 or 2 experimental points (EPs), worth $0.25 each. Two EPs were endowed for every payoff shot. If a bet was not placed within the 10-second interval it was not eligible for payment and was not used in the data analysis. After the freeze-frame, the tape rolled and the outcome of the shot was displayed.

The first 3 quarters of the Feb 3, 2008 game vs. the Washington Wizards and first half of the Feb 8, 2008 game vs. the Orlando Magic were used in the study.[2] Subjects were told in advance that 2 games were being used and were given a five minute break between games. The tape only included the Lakers' shots so more shots could be displayed in the time frame. It was played without sound so that the behavior of subject's could not be influenced by the commentator's opinions.[3] For all shots, the shooter's name was prominently displayed on screen so that subjects could easily identify the player. For payoff shots the shooter's name remained on screen for the entire freeze-interval. There were 37 payoff shots out of a total of 107 shots. The tape duration was 37.5 minutes.

The challenge of using a subjective probability sequence is formulating payoffs so that beliefs concerning serial dependence can be inferred. NBA players take shots of widely varying difficulty. For example, 40% is considered a very good shooting percentage for a long-range shot while an open shot from 10 feet is made over 70% of the time. If payoffs were constant across shot difficulty, there would be little variation in bets based on perception of autocorrelation because shot difficulty (or perceptions thereof) would drive betting behavior.

The solution was to weight payoffs by difficulty. This was possible because the game tape used came from a larger sample, collected by the author and employed in other papers, of 60 games from the 2007/08 Laker's season consisting of over 5000 shots.[4] The data set was built through game logs and by actually watching the games to record all the relevant shot conditions. The shot conditions (distance, defensive pressure, shot type, time on the shot clock, physical location) were used to predict success rates ($\hat{p}$) for the shots used in the experiment. The data set and regressions used to formulate the success rates is described in more detail in Rao (2009a) and Rao (2009b).

Predicted success rate $\hat{p}$ is taken as the measure of shot difficulty. Accordingly, a 1 point bet on hit won $\frac{1}{\hat{p}}$ if correct and 0 otherwise. A single bet on miss earned $\frac{1}{1-\hat{p}}$ if correct and 0

[2] The experimental game tape is available from the author, but can only be used for private viewing due to copyright restrictions.

[3] Basketball commentators frequently refer to players as "heating up" or "on fire."

[4] The data set was also used to support the result cited earlier that NBA players exhibit a belief in the hot hand through taking sub-optimally difficult shots following a string of hits.

otherwise.[5] Notice that this means that if one felt the model predictions (based on observable shot conditions) were correct, she would be indifferent between "hit" and "miss" as both offer the same expected value in this case. The experimenter carefully explained how payoffs were constructed and informed the subjects that it was in their financial interest to bet "hit" when they felt the player was more likely to make the shot than predicted by the model which used *only observable shot conditions* and bet "miss" when they felt the player was less likely to make the shot than predicted by the model. This point was reinforced by the notion that they goal was to "beat the model" and they had a financial incentive to do so. The shot conditions used were read to the subjects twice, specifically these did not include past shot outcomes, however the experimenter did not draw attention to this omission for fear of tainting the sample.[6]

Without the quantification of success probability, it would be difficult, if not impossible to study GF and HHF type beliefs with a subjective sequence such as professional basketball shots. Payoffs were designed to mimic those of purchasing or selling stock on an efficient exchange. Notice that in both cases, in the absence of private information both possible actions (buy or short, bet "hit" or "miss") offer the same expected value. In such an environment beliefs about autocorrelation are likely to affect observed behavior.

Importantly, subject's were not told that the study was examining beliefs about serial dependence – the words "hot hand" or "gambler's fallacy" were not used in the instructions. An instructional video was shown with three example shots and displayed payoffs for each bet. It gave the subjects a feel for the length of the freeze-interval and added credibility to the design. Full instructions are in the appendix.

4.3 Results

The main method of analysis is to examine how betting behavior changes as a function of a player's past performance. The debate over whether the hot hand truly exists in professional basketball has raged since Gilovitch, Villone and Tversky (1985) famously proclaimed that it does not. The result however was criticized because the analysis did not control for shot difficulty. Rao (2009a), using the data which generated $\hat{p}$ for this study, shows that controlling for difficulty there is no evidence of increased ability after makes or decreased ability after misses. The paper not only finds no evidence for the hot hand it also bounds possible hot hand effects as being very minimal, if they exist at all.[7]

Table 1 examines how betting behavior changes with past performance. The dependent

[5] $\hat{p}$ was not displayed on screen. The purpose of the design was to pick up beliefs in autocorrelation not to see if say, subjects were correct/incorrect about the difficulty of certain shots. However, $\hat{p}$ was displayed on the instructional tape. This was done to add credibility to the design and to help explain payoffs.

[6] No subjects asked about the omission of past shot outcomes. All questions were answered privately.

[7] The analysis cannot reject the existence of very rare departures from a stationary model. So if a player is thought to get "hot" a few times per season, this cannot be rejected by the data. What is rejected is a systematic dependence on past performance.

Table 4.1: Fixed-Effect OLS Regressions, Effect of Streaks on Betting Behavior

Dependent Variable	1{Bet="Hit"}			Wager Amount		
Sample	All	Experts	Novices	All	Experts	Novices
Player made	-0.0455	-0.0367	-0.0567	0.0150	0.0249	0.00262
last shot	(0.033)	(0.043)	(0.052)	(0.036)	(0.049)	(0.052)
Player made	-0.0735**	-0.0320	-0.126**	-0.124***	-0.188***	-0.0437
last 2 shots	(0.033)	(0.043)	(0.051)	(0.035)	(0.048)	(0.051)
Player made	0.0506*	0.0178	0.0918**	0.0729**	0.137***	-0.00780
last 3 shots	(0.030)	(0.040)	(0.047)	(0.033)	(0.044)	(0.047)
Player missed	-0.103***	-0.134***	-0.0626	-0.0363	-0.121**	0.0707
last 2 shots	(0.037)	(0.048)	(0.057)	(0.039)	(0.054)	(0.057)
Cubic $\hat{p}$ control	X	X	X	X	X	X
Subjects	70	39	31	70	39	31
Obs	2520	1404	1116	2520	1404	1116

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

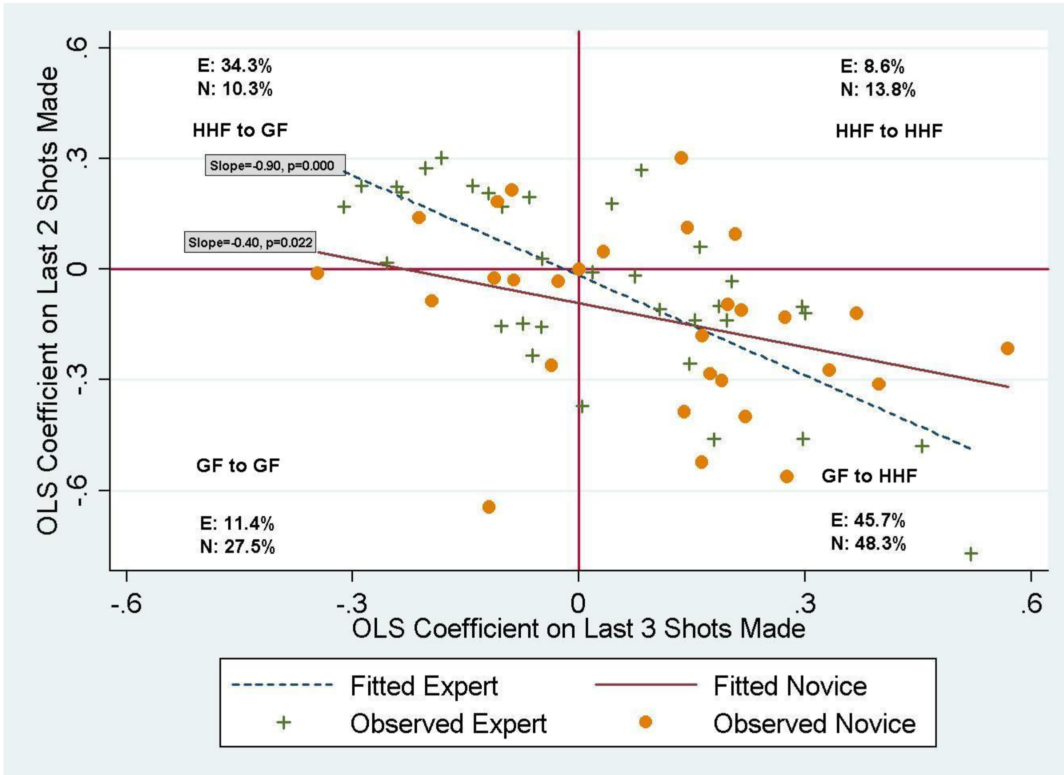
variable in columns 1-3 is bet outcome ("hit" or "miss") and the dependent variable in columns 4-6 is wager amount (1 or 2 EPs). The table supports the main claim of the paper GF to HHF transition. Subject fixed-effects and shot difficulty controls (cubic in $\hat{p}$) serves to reduce the influence of risk aversion.[8]

We see that novice subjects were significantly more likely to bet "miss" following 2 made shots by the player (GF). The causal effect of the third shot made was a significant increase in the probability of betting "hit" (HHF). However it must be noted that since GF beliefs dominate for short sequences, the hot hand belief serves as a correction. In aggregate, expert subjects showed the same pattern with respect to string of hits, but with lower magnitudes and lacking statistical significance. They show significant HHF after 2 straight misses by the player; novice subjects do not exhibit this pattern. The effect of a single shot made or missed was indistinguishable from zero for both groups.

The idea behind giving subjects a choice in wager amount was to infer strength of belief. Table 1 shows that novice subjects did not change their betting behavior systematically with past outcomes. It was not, as one might suspect, simply that they were always wagering 1 EP; the mean for both groups was similar (1.54 experts, 1.52 for novices). Experts wagered significantly more after a player's string of 3 makes. Initially this is puzzling since column (2) shows they show little pattern in their shot predictions. The reason is given in Figure 1, which is explained in detail below. While on net the expert subjects did not show a significant propensity to bet "hit" after 3 makes, there is a group of subjects with significantly positive coefficients and a group with significantly negative coefficients. The resultant aggregate estimate is near zero and

[8]Rao (2009a) finds that players take harder shots when they are on a make streak. If subject's had an unwillingness to bet hit for hard shots (risk aversion) this would work against finding hot hand effects. Including a flexible difficulty solves this problem.

Figure 4.1: Subject-by-subject analysis



belies the fact that beliefs are indeed changing.

The regressions in columns (2) and (3) were run subject-by-subject to estimate individual coefficients. Figure 1 shows the relationship between the coefficient on "player made last 2 shots" and "player made last 3 shots." The percentages of subjects falling in each quadrant support result presented in Table 1 as the most common beliefs are GF after 2 makes and HHF after 3 makes (the lower right quadrant). The upper right quadrant, representing the classical hot hand belief, is virtually empty.

The figure also plots OLS linear fits, which reveal a statistically significant negative relationship between holding GF and HHF beliefs ($p = 0.0000$ for experts and $p = 0.022$ for novices). If the subject displayed GF for streak length 2 then HHF is predicted for streak length 3. If HHF is observed for streak length 2 then GF is predicted for streak length 3. The former transition is precisely the prediction of the Rabin (2002) model and this is the modal pattern. Furthermore, the relative magnitudes of the bias are positively related (another feature of the model). Contrary to the model's predictions a significant fraction of transitions are HHF to GF. That is, the relationship between the two inference mistakes is strong regardless of the initial bias. This an interesting finding as there do not appear to be GF-types or HHF-types, rather subjects flip from one bias to the other depending on the circumstance.

Had I found a null result (i.e. no effect of streaks) one could argue that keeping track of the many concurrent player streaks is too difficult a task for experimental subjects. In fact this criticism would be even stronger for novice subjects as they are unfamiliar with the players and the speed of professional basketball. However, even though these novice subjects were never told monitoring past player performance was important, their shot predictions indicate they naturally did so and responded in systematic ways.

In the post questionnaire, subjects were asked about their belief in the hot hand and gambler's fallacy. 32 (of 39) experts and 22 (of 31) novices reported that they believed a player is more likely than normal to make a shot following a made shot and streak of 3 makes (they reported similar beliefs after a single made shot). 19 experts and 14 novices reported they would bet black in roulette after a streak of reds (only 2 of each reported they would stay with red, the rest were indifferent).

78% subjects reported a belief in the hot hand for basketball while only 7% report a belief in a "gambler's hand" (15% said past outcomes did not matter). However, these reported beliefs do not show up in betting behavior. Appendix Table 1 presents regressions of the OLS coefficients from Figure 1 on the self-reported GF and HHF beliefs. For experts, the reported beliefs are totally uncorrelated with their bets. For novices, a reported HHF belief predicts a significantly higher coefficient on "last 2 made" and a lower coefficient on "last 3 made." The sign switch is not surprising in light of Figure 1. The results lend further credence to the argument that treating these biases as unchanging "types" and surveying people to determine their type is a flawed methodology.

In an interesting contrast to the beliefs of observers, basketball players themselves appear to only believe in the hot hand, if at all. Rao (2009a) shows that players who respond to their past performance (half the sample) do so by taking more difficult shots, passing the ball less and shooting more frequently. The response to streaks of 3 or 4 makes is generally stronger than the response to a streaks of 1 or 2 makes, but the direction is the same. That is, there is no evidence of switching into easier shots and shooting less frequently after short sequences of makes. This suggests a dichotomy between the beliefs of observers of a sequence and the practitioners themselves – interested readers are directed to Rao (2009a).[9]

One reason for the great interest in both GF and HHF is that it can help explain departures from rationality in the financial markets. The disposition effect, the tendency to lock in small gains in equity positions but not close out small losses, can be explained by GF beliefs.[10] HHF beliefs can potentially drive bubbles, as stocks get "hot" they increase further beyond levels driven by fundamentals.

[9]Understanding the players' beliefs likely involves an account of the physiological feedback of a make versus a miss. A discussion of why players believe in the hot hand when there is no evidence it exists is beyond the scope of this paper.

[10]On an efficient exchange, selling at any point is rational in the absence of transactions costs. However, for many investors locking in small gains puts a drag on their net returns due to the accumulation of transactions costs.

In light of these observations, I examine betting behavior as a function of the success and failure of past bets in Table 2. Notice that the unit of observation has changed and is now beliefs about how their bet returns are correlated over time.[11] After a string of correct bets, a "personal hot hand" belief dictates sticking to the current bet and increasing the wager amount; a personal "gambler's hand" belief would be revealed through a pattern of switching the bet and reducing the bet number. For strings of past bet failures, switching indicates a "cold hand" belief and sticking with current bet can interpreted as "my luck is bound to change." Roughly, these correspond to HHF and GF respectively, although some authors posit that HHF refers a belief in positive autocorrelation following success only (i.e. a belief in the hot hand does not imply a belief in a cold hand).

Column (1) examines subjects' propensity to keep their bet unchanged as a function of past betting performance. The coefficients on the past success dummies are all negative, indicating switching after past success. Individual significance levels are confounded by the co-linearity of the regressors, but a Wald test rejects that they are jointly zero ($p = 0.049$, two-tailed test). There was significantly more switching after 3 incorrect bets (HHF) and significantly less after 4 incorrect bets (GF). This provides further evidence of the interconnectedness of the two biases and is our first example of a HHF to GF transition in aggregate. Column (2) shows that there is a slight pattern of decreasing wager amount after past success and increasing after past failures, however none of the coefficients are statistically significant and they have a small magnitude.[12]

It is interesting to note that in both columns, streaks of length 2 or shorter had very little effect – all coefficients are estimated to be within 0.005 of zero. Juxtaposed to the findings of Table 1 this result shows that the effect of streak length is situationally dependent. Rabin (2002) models a belief in the law of small numbers using an agent who perceives random outcomes as draws *without replacement* from an urn of size N. The size of the hypothetical urn determines how "unusual" the agent views streaks. With a large urn many streaks naturally occur, while with a small urn oscillation is expected to kick in quite quickly. In the language of the Rabin model, the urn appears bigger for inference of personal autocorrelation as compared to autocorrelation of the professional basketball players.

4.4 Conclusion

The hot hand fallacy and gambler's fallacy are seemingly opposing biases in the perception autocorrelation. HHF says that people over-estimate autocorrelation while GF says they

[11] It is also natural to examine the effect of "team streaks." Regressions available from the author show that team streaks had little effect (slight evidence of GF). For this reason and due to space considerations, the analysis of team streaks is not included in the paper.

[12] In contrast, Croson and Sundali (2005) find that casino gambler's display personal hot hand effects with bets in craps and roulette.

Table 4.2: Fixed-Effect OLS Regressions: Effect of Past Bet Outcomes on Current Round Betting Behavior

Dependent Variable	1{Bet=Last Bet}	Wager Amount
1{Last bet correct}	-0.00567	-0.00519
	(0.0247)	(0.0237)
1{Last 2 bets correct}	-0.00291	-0.00469
	(0.0301)	(0.0289)
1{Last 3 bets correct}	-0.0391	-0.0553
	(0.0429)	(0.0412)
1{Last 4 bets correct}	-0.0729	-0.0122
	(0.0471)	(0.0452)
1{Last 2 bets incorrect}	0.0125	0.0247
	(0.0336)	(0.0323)
1{Last 3 bets incorrect}	-0.110*	0.0350
	(0.0602)	(0.0578)
1{Last 4 bets incorrect}	0.232**	-0.115
	(0.0930)	(0.0893)
Cubic $\hat{p}$ control	X	X
Subjects	39	31
Obs	1404	1116

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

underestimate it. Yet previous research has found strong evidence for both. To reconcile this apparent inconsistency, authors have argued that both biases have a common root cause in a belief in the law of small numbers. Rabin (2002) presents a formal model of this argument.

If the mistakes have a common cause, then not only should we expect both to manifest themselves in experimental and field data, but also they should coexist within a single agent. I find strong support for this (perhaps initially perplexing) assertion. Most subjects in the study display GF for short streaks and HHF for long streaks. While the GF to HHF transition predicted by the Rabin model is most common, there is a significant portion of subjects who display precisely the inverse pattern. Somewhat surprisingly, even though the sequence used (basketball shot outcomes) is thought to be the hot hand's domain, less than 10% of subjects showed consistent hot hand beliefs.

The results of this paper can be applied to decisions in financial markets and everyday choices alike. Indeed most probability sequences people encounter are likely more similar to the subjective world of basketball shooting than the cold objective world of roulette wheels, dice and urns that have been the standard vehicle to analyze inference mistakes. The payoffs used in the study duplicate those of returns on an efficient exchange. Using this realistic environment, the paper finds that GF and HHF can peacefully coexist within a single agent.

Table 4.3: Correlation of Post-Questionaire GF and HHF Questions with Observed Behavior

Dependent Variable	β Last 2 Made	β Last 2 Made	β Last 3 Made	β Last 3 Made
Sample	Expert	Novice	Expert	Novice
1{Reported GF Belief}	-0.0744	0.00113	0.0718	0.00114
	(0.0796)	(0.0814)	(0.0654)	(0.0665)
1{Reported HHF Belief}	0.000260	0.200**	-0.0222	-0.113
	(0.0600)	(0.0726)	(0.0524)	(0.0902)
Constant	0.00401	-0.269***	0.00106	0.174*
	(0.0543)	(0.0624)	(0.0451)	(0.0977)
Subjects	39	30	39	30

*** p<0.01, ** p<0.05, * p<0.1

Robust standard errors in parentheses

4.5 Appendix

4.5.1 Tables

Table 4.4 Game Tape Shot Details

Player	**Possession**	**Hit**	$\hat{p}$	**Payoff Shot**
Kobe Bryant	36	0	0.22	1
Ronny Turiaf	37	1	0.09	0
Kobe Bryant	38	1	0.56	0
Ronny Turiaf	39	0	0.36	0
Kobe Bryant	40	0	0.12	0
Derek Fisher	41	1	0.5	1
Ronny Turiaf	42	1	0.39	0
Vladimir Radmanovic	43	1	0.44	1
Ronny Turiaf	44	0	0.15	0
Kobe Bryant	45	1	0.31	0
Kobe Bryant	46	0	0.22	1
Vladimir Radmanovic	47	1	0.43	1
Ronny Turiaf	48	0	0.24	0
Ronny Turiaf	49	0	0.44	0
Ronny Turiaf	50	1	0.94	0
Ronny Turiaf	51	1	0.36	0
Vladimir Radmanovic	52	0	0.45	1
Derek Fisher	53	0	0.38	0
Lamar Odom	54	1	0.96	0
Vladimir Radmanovic	55	1	0.43	1
Vladimir Radmanovic	56	0	0.51	1
Ronny Turiaf	57	0	0.17	0
Lamar Odom	58	1	0.27	1
Sasha Vujacic	59	0	0.47	0
Jordan Farmar	61	0	0.3	1
Jordan Farmar	62	0	0.31	1
Sasha Vujacic	63	1	0.46	1

Table 4.4: Game Tape Shot Details
Feb 3, 2008 vs. Washington Wizards

Player	**Possession**	**Hit**	$\hat{p}$	**Payoff Shot**
Lamar Odom	1	0	0.29	0
Kobe Bryant	2	0	0.17	0
Kobe Bryant	3	1	0.37	0
Kobe Bryant	4	0	0.49	1
Ronny Turiaf	5	1	0.48	0
Derek Fisher	6	0	0.38	0
Vladimir Radmanovic	7	1	0.92	0
Kobe Bryant	8	1	0.63	0
Derek Fisher	9	1	0.56	1
Kobe Bryant	10	0	0.17	0
Lamar Odom	11	0	0.56	0
Ronny Turiaf	12	0	0.24	0
Kobe Bryant	13	1	0.63	0
Kobe Bryant	14	0	0.38	1
Vladimir Radmanovic	15	1	0.14	1
Kobe Bryant	16	0	0.15	0
Lamar Odom	17	1	0.46	1
Derek Fisher	18	0	0.52	0
Kobe Bryant	19	1	0.34	1
Kobe Bryant	20	1	0.22	1
Kobe Bryant	21	1	0.91	0
Sasha Vujacic	22	0	0.23	1
Jordan Farmar	23	0	0.08	0
Lamar Odom	24	0	0.25	0
Lamar Odom	25	0	0.29	0
Lamar Odom	27	1	0.5	1
Lamar Odom	28	1	0.53	0
Jordan Farmar	29	0	0.21	0
Luke Walton	30	1	0.4	0
Lamar Odom	31	1	0.29	1
Jordan Farmar	32	1	0.23	0
Sasha Vujacic	33	1	0.53	1
Vladimir Radmanovic	34	1	0.45	1
Kobe Bryant	35	1	0.38	1

Table 4.4 (cont.) Feb 8, 2008 vs. Orlando Magic

Player	**Possession**	**Hit**	$\hat{p}$	**Payoff Shot**
Kobe Bryant	1	0	0.31	1
Kobe Bryant	2	1	0.33	0
Kobe Bryant	3	1	0.7	1
Pau Gasol	4	1	0.44	0
Pau Gasol	5	0	0.27	0
Kobe Bryant	6	0	0.38	1
Vladimir Radmanovic	7	1	0.94	0
Kobe Bryant	8	0	0.38	0
Kobe Bryant	9	0	0.58	0
Kobe Bryant	10	1	0.91	0
Kobe Bryant	11	0	0.14	0
Derek Fisher	12	0	0.36	0
Pau Gasol	13	1	0.74	0
Kobe Bryant	14	0	0.39	1
Derek Fisher	15	0	0.19	0
Vladimir Radmanovic	16	1	0.43	1
Kobe Bryant	17	1	0.91	0
Vladimir Radmanovic	18	1	0.42	1
Lamar Odom	19	1	0.19	0
Pau Gasol	20	1	0.52	0
Vladimir Radmanovic	21	0	0.44	1
Kobe Bryant	22	1	0.39	0
Jordan Farmar	23	0	0.38	0
Kobe Bryant	24	1	0.55	0
Kobe Bryant	25	0	0.39	1
Ronny Turiaf	26	0	0.36	0
Kobe Bryant	27	0	0.32	0
Kobe Bryant	28	0	0.13	0
Kobe Bryant	29	1	0.93	0
Sasha Vujacic	30	1	0.5	1
Jordan Farmar	31	0	0.31	0
Ronny Turiaf	32	0	0.49	0
Sasha Vujacic	33	0	0.23	0
Pau Gasol	34	1	0.56	0
Kobe Bryant	35	0	0.26	0
Lamar Odom	36	1	0.65	0
Sasha Vujacic	37	1	0.47	1
Kobe Bryant	38	0	0.38	1
Sasha Vujacic	39	0	0.52	0
Pau Gasol	40	0	0.41	1
Kobe Bryant	41	0	0.26	0
Kobe Bryant	42	0	0.39	0
Pau Gasol	43	0	0.25	0
Pau Gasol	44	1	0.25	0
Sasha Vujacic	45	1	0.47	0

4.5.2 Experimental Instructions

Welcome

Thank your for participating in our experiment. We will begin shortly. Today's experiment will last about 1 hour. Although earnings will vary by subject, most subjects will earn between $10 and $25.

Informed Consent

Placed in front of you is an informed consent form to protect your rights as a subject. Please read and sign it. If you would like to choose not to participate in the study you are free to leave at this point. If you have any questions, we can address those now.

Anonymity

Your anonymity in this study is assured. Your name will never be collected or connected to any decision you make here today. Your email address was collected for invitation purposes only and will never be connected to your decisions. Furthermore, your earnings will be paid in a sealed envelope with your subject number so that even those running the study will not know your earnings. No other subject in the study will know any of your choices or performance in the study.

Rules

- Quiet please, please do not talk or communicate with other subjects.
- Please turn your cell phones off.
- If you have a question at any point, just raise your hand.
- Please put away any books, papers, computers, etc. that you have brought with you.

Your Earnings

The decisions you make today will determine your earnings. We will explain exactly how earnings will be calculated at the appropriate time. Your earnings will be paid in cash, placed in a sealed envelope with your subject number. Although earnings will vary by subject, most subjects will earn between $10 and $25.

Today's Experiment

In this experiment, you will be asked to predict the outcomes of shots taken in NBA basketball games. You were chosen for the experiment because you indicated on your entry into the subject database that you are a basketball fan. During the experiment you will be shown 5 quarters of basketball from the 2007/08 Lakers' season (i.e. last season). Specifically you will be watching the first 3 quarters of one game and the first half of another. The tape has been edited so only the Lakers' shots are shown. While the opponents shots are edited out, none of the Lakers' shots are skipped. The tape will be played without sound. The shooters' names are displayed on screen to help you identify the player shooting the ball. The tape duration is 36 minutes.

The game will be projected on the screen in front of you.

For a number of shots, called "payoff shots", the tape is freeze-framed just as the ball leaves the shooter's hand. The freeze-frame will remain on the screen for 10 seconds. During this time you will be asked to provide your prediction of the shot outcome. The exact method of your payment will be explained shortly – it will depend directly on the accuracy of your predictions. At the end of the instructions an example tape of three shots will be played to show you exactly how things work.

During the experiment you will make predictions for 43 payoff shots. For each payoff shot you will be endowed 2 experimental points (EPs), each valued at $0.20. For each payoff shot you are required to "bet" at least 1 and at most 2 of the endowed EPs. We will explain shortly the payoffs for these bets and how they are actually "placed". At the end of the experiment your EPs will be converted to cash and you will be paid in an envelope with your subject number.

The payoffs for your shot prediction bets are a bit complicated, so we will go through them slowly.

For the 2007/08 Lakers' season, a data set of 60 games (over 5000 shots) was compiled by watching the games and recording shot distance, location, defensive pressure, shot type (i.e. turnaround, fadeaway, hook) and if the shot was forced due to less than 5 seconds on the shot clock. Every effort was made to record all relevant shot conditions. These variables, , where used to predict the shooting percentage for every shot in the sample. The predictions were made by player, to allow for differences in ability and shot preference. Specifically the the model used to predict shot success was a popular binary dependent variable routine called "Probit."

The econometric model provides an estimated success rate for every shot taken on the edited game tape you will view. For each shot, we will call this estimated success rate $\hat{p}$ – this is a probability ranging from 0 (no chance of making) to 1 (made with certainty) and it can take any value in between. For example if $\hat{p}$=0.5, this means that the model predicts that 50% of the time the shot will be made, while 50% of the time the shot will be missed.

In front of you are 37 numbered "Bet Cards" each corresponding to a payoff shot. On them there are two words "hit" and "miss" and a blank by the word "Bets". For each shot you circle the "hit" or "miss" to make your bet and write the number of bets you would like to place, 1 or 2, in the blank beside the word bets.

The monetary incentives in this study are designed so that you should bet "hit" when you feel the shot is more likely to go in *than predicted by the model* which uses *observable shot conditions alone* and "miss" when you feel the shot is *less likely to go in based on the model's predictions*. To those ends, the payouts are as follows: for every EP bet on "hit" you will earn $\frac{1}{\hat{p}}$ EPs if the shot goes in and 0 EPs if the shot is a miss; for every EP you bet on miss you will earn $\frac{1}{1-\hat{p}}$ EPs if the shot is in fact a miss and 0 EPs if the shot goes in. In a sense, your goal is to "beat the model" and these payoffs ensure that you have a financial incentive to try and do so.

Here is an example: suppose the payoff shot has a predicted success rate of 0.40 (40%). In this case a 1 point bet on "hit" earns 1/.4=2.5 points if the shot goes in and 0 points if the shot is a miss; a 1 point bet on "miss" earns 1/(1-.4)=1.67 points if the shot is a miss and 0 points if the shot goes in. Notice that in this example the outcome that the model predicts is less likely, hit, offers a higher payoff. If you felt that the model's predictions were exactly right then both bets have the same expected value: .4*(1/.4)+.6*0=1 for a one point bet on hit and .6*(1/.6)+.4*0=1 for a one point bet on miss. If you think "hit" is more likely than the model predicts (i.e. above 40%) then a bet on hit offers expected value in excess of 1 point whereas a bet on "miss" offers expected value less than 1. As such, in that hypothetical case you should bet on hit to make the most money on average. Conversely if you think the true success probability is less than the model's predictions for the shot, you are best served betting "miss."

In front of you there is as stack of note cards, numbered 1-43 corresponding to the payoff shots that will be displayed in that order. When you a see freeze-frame, you must place your bet for that shot within the 10 second "freeze interval". To do so, circle either "hit" or "miss" and write down the number of bets, 1 or 2, you would like to place on that shot. You must then place the card in the box placed in front of you. To be eligible for payment, the card must be in the box before the 10 second freeze frame ends and the shot outcome is revealed. Please place the cards face down in the box, as this will help us when we compute payments at the end. We will be passing through the aisles to ensure that all bets are placed within the allotted time.

In a few moments we will watch the sample game tape and answer questions. To recap:

1. For every payoff shot you will be endowed 2 EPs, you must bet at least 1 of these points.
2. If you bet on "hit" you will earn $\frac{1}{\hat{p}}$ points if the shot goes in and 0 points if the shot is a miss
3. If you bet on "miss" you will earn $\frac{1}{1-\hat{p}}$ points if the shot is a miss and 0 points if the shot is a hit
4. $\hat{p}$ is the predicted success rate for the shot based only on the shot's observable conditions, these were formulated used a data set of 5,000 shots
5. You must place your bets during the 10 second freeze frame window, after this interval the tape will roll and the outcome of the shot will be shown
6. The payoffs are designed so that you will make the most money on average by betting **"hit" when you think the shot is more likely to go in than predicted by the model** and **"miss" when you think the shot is less likely to go in than predicted by the model.**
7. Another way of putting this is that your job is to "beat the model."

The following video shows three sample payoff shots, after the shots the predicted success probabilities are shown and the amount of points that would have been earned for a bet on each outcome are shown.

[Video shown]

Are there any questions? The game tape will now begin, it will last 36 minutes. After this your payments will be computed and the experiment will conclude. Please try to pay close attention throughout the tape. The validity of our research depends on your honest efforts, as does your monetary payoff.

Bibliography

[Alicke and Govorun, 2005] Alicke, M. and Govorun, O. (2005). *The Better-than-Average Effect*. Psychology Pr.

[Andreoni and Bernheim, 2009] Andreoni, J. and Bernheim, B. (2009). Social Image and the 50-50 norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica*, 77:1607–1636.

[Babcock and Loewenstein, 1997] Babcock, L. and Loewenstein, G. (1997). Explaining Bargaining Impasse: The Role of Self-serving Biases. *The Journal of Economic Perspectives*, pages 109–126.

[Babcock et al., 1995] Babcock, L., Loewenstein, G., Issacharoff, S., and Camerer, C. (1995). Biased Judgments of Fairness in Bargaining. *The American Economic Review*, pages 1337–1343.

[Babcock et al., 1996] Babcock, L., Wang, X., and Loewenstein, G. (1996). Choosing the Wrong Pond: Social Comparisons in Negotiations that Reflect a Self-serving Bias. *The Quarterly Journal of Economics*, 111(1):1–19.

[Barber and Odean, 2001] Barber, B. and Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, 116(1):261–292.

[Becker et al., 1964] Becker, G., DeGroot, M., and Marschak, J. (1964). Measuring Utility by a Single-response Sequential Method. *Behavioral Science*, 9(3):226–232.

[Bernard and Thomas, 1989] Bernard, V. and Thomas, J. (1989). Post-earnings-announcement Drift: Delayed Price Response or Risk Premium? *Journal of Accounting research*, 27:1–36.

[Bhattacharya et al., 2009] Bhattacharya, J., Goldman, D., and Sood, N. (2009). Market Evidence of Misperceived Mortality Risk. *Journal of Economic Behavior and Organization*.

[Brunnermeier and Parker, 2005] Brunnermeier, M. and Parker, J. (2005). Optimal Expectations. *The American Economic Review*, 95(4):1092–1118.

[Burns, 2004] Burns, B. (2004). Heuristics as beliefs and as behaviors: The adaptiveness of the hot hand. *Cognitive Psychology*, 48:295–331.

[Burns and Corpus, 2004] Burns, B. and Corpus, B. (2004). Randomness and induction from streaks: Gamblers fallacy vs. hot hand. *Psychonomic Bulletin and Review*, 11:179–184.

[Camerer, 1989] Camerer, C. F. (1989). Does the basketball market believe in the 'hot hand,'? *The American Economic Review*, 79(5):1257–1261.

[Carhart, 1997] Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82.

[Carrillo and Mariotti, 2000] Carrillo, J. and Mariotti, T. (2000). Strategic Ignorance as a Self-disciplining Device. *Review of economic studies*, pages 529–544.

[Chiappori et al., 2002] Chiappori, P.-A., Levitt, S., and Groseclose, T. (2002). Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *The American Economic Review*, 92(4):1138–1151.

[Christensen-Szalanski and Bushyhead, 1988] Christensen-Szalanski, J. and Bushyhead, J. (1988). Physicians' Use of Probabilistic Information in a Real Clinical Setting. *Professional Judgment: A Reader in Clinical Decision Making*.

[Clotfelter and Cook, 1993] Clotfelter, C. and Cook, P. (1993). The "Gambler's Fallacy" in Lottery Play. *Management Science*, pages 1521–1525.

[Compte and Postlewaite, 2004] Compte, O. and Postlewaite, A. (2004). Confidence-enhanced Performance. *American Economic Review*, pages 1536–1557.

[Croson and Sundali, 2005] Croson, R. and Sundali, J. (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty*, 30(3):195–209.

[Darley and Gross, 1983] Darley, J. and Gross, P. (1983). A hypothesis confirming bias in labeling effects. *Journal of Personality and Social Psychology*, XLIV.

[Eidinger and Schapira, 1984] Eidinger, R. and Schapira, D. (1984). Cancer Patients' Insight Into Their Treatment, Prognosis, and Unconventional Therapies. *Cancer*, 53(12).

[Eil and Rao, 2009] Eil, D. and Rao, J. M. (2009). The good news-bad news effect: Assymetric processing of objective information about yourself. *UCSD*, Working paper.

[Erev and Roth, 1998] Erev, I. and Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review*, 88(4):848–881.

[Estes, 1950] Estes, W. (1950). Toward a Statistical Theory of Learning. *Psychological Review*, 57(2):94–107.

[Fehr and List, 2004] Fehr, E. and List, J. (2004). The Hidden Costs and Returns of Incentives: Trust and Trustworthiness among CEOs. *Journal of the European Economic Association*, 2(5):743–771.

[Festinger, 1957] Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press: Stanford, CA.

[Fisman et al., 2006] Fisman, R., Iyengar, S., Kamenica, E., and Simonson, I. (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *Quarterly Journal of Economics*, 121(2):673–697.

[Foster et al., 1984] Foster, G., Olsen, C., and Shevlin, T. (1984). Earnings Releases, Anomalies, and the Behavior of Security Returns. *The Accounting Review*, 59(4):574–603.

[Frey, 1981] Frey, D. (1981). The Effect of Negative Feedback About Oneself and Cost of Information on Preferences for Information About the Source of this Feedback. *Journal of Experimental Social Psychology*, 17(1):42–50.

[Gelman, 2007] Gelman, A. (2007). Letter to the editors regarding some papers of Dr. Satoshi Kanazawa. *Journal of Theoretical Biology*, 245(3):597–599.

[Gilovitch et al., 1985] Gilovitch, T., Villone, R., and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17:295–314.

[Grether, 1980] Grether, D. (1980). Bayes Rule as a Descriptive Model: The Representativeness Heuristic. *Quarterly Journal of Economics*, 95:537–557.

[Heider, 1958] Heider, F. (1958). *The Psychology of Interpersonal Relations*. Wiley: New York.

[Holt, 1986] Holt, C. (1986). Scoring-rule Procedures for Eliciting Subjective Probability and Utility Functions. *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti. North Holland Press, Amsterdam.*

[Kanazawa, 2007] Kanazawa, S. (2007). Beautiful parents have more daughters: A further implication of the generalized trivers–willard hypothesis. *Journal of theoretical biology*, 244(1):133–140.

[Kareev, 1995] Kareev, Y. (1995). Positive bias in the perception of covariation. *Psychological Review*, 102(3).

[Kelley and Michela, 1980] Kelley, H. and Michela, J. (1980). Attribution Theory and Research. *Annual review of psychology*, 31(1):457–501.

[Köszegi, 2006] Köszegi, B. (2006). Ego utility, Overconfidence, and Task Choice. *Journal of the European Economic Association*, 4(4):673–707.

[Larwood, 1978] Larwood, L. (1978). Swine flu: A Field Study of Self-serving Biases. *Journal of Applied Social Psychology*, 8(3):283–289.

[Lord et al., 1979] Lord, C., Ross, G., and Lepper, L. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Personality and Social Psychology*, 37(11).

[Mahoney, 1977] Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1(2).

[Massey and Thaler, 2007] Massey, C. and Thaler, R. (2007). The loser's curse: Over confidence vs. market efficiency in the nfl draft. SSRN Workping Paper.

[McCabe et al., 2000] McCabe, K., Mukherji, A., and Runkle, D. (2000). An experimental study of information and mixed-strategy play in the three-person matching-pennies game. *Economic Theory*, 15(2):421–462.

[Messick and Sentis, 1979] Messick, D. and Sentis, K. (1979). Fairness and Preference. *Journal of Experimental Social Psychology*, 15(4):418–434.

[Metrick, 1995] Metrick, A. (1995). A natural experiment in" Jeopardy!". *The American Economic Review*, 85(1):240–253.

[Mischel, 1976] Mischel, W. (1976). Determinants of selective memory about the self. *Journal of Consulting and Clinical Psychology*, 44(1):92–102.

[Moore and Healy, 2008] Moore, D. and Healy, P. (2008). The Trouble with Overconfidence. *Psychological Review*, 115(2):502.

[Nickerson, 1998] Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2):175–220.

[Ochs, 1995] Ochs, J. (1995). Games with unique, mixed strategy equilibria: An experimental study. *Games and Economic Behavior*, 10(1):202–217.

[O'Neill, 1987] O'Neill, B. (1987). Nonmetric test of the minimax theory of two-person zerosum games. *Proceedings of the National Academy of Sciences*, 84(7):2106–2109.

[Pakes, 1986] Pakes, A. (1986). Patents as options: Some estimates of the value of holding european patent stocks. *Econometrica*, 54(4):755–784.

[Palacios-Huerta, 2003] Palacios-Huerta, I. (April 2003). Professionals play minimax. *The Review of Economic Studies*, 70:395–415(21).

[Palacios-Huerta and Volij, 2008] Palacios-Huerta, I. and Volij, O. (2008). Field centipedes. *Available at: http://www.hhs.se/NR/rdonlyres/46CE89B9-1E92-4829-A305-3272D5C52411/0/080509Huerta.pdf*.

[Paling, 2003] Paling, J. (2003). Strategies to Help Patients Understand Risks. *British Medical Journal*, 327(7417):745–748.

[Plous, 1991] Plous, S. (1991). Biases in the assimilation of technological breakdowns: Do accidents make us safer? *Journal of Applied Social Psychology*, 21.

[Provencher, 1997] Provencher, B. (1997). Structural versus reduced-form estimation of optimal stopping problems. *American Journal of Agricultural Economics*, 79(2):357–368.

[Rabin and Schrag, 1999] Rabin, M. and Schrag, J. (1999). First impressions matter: A model of confirmatory bias. *Quarterly Journal of Economics*, 114(1):37–82.

[Rao, 009aa] Rao, J. M. (2009aa). Experts' perceptions of autocorrelation: The hot hand fallacy among professional basketball players. UCSD.

[Rao, 009ab] Rao, J. M. (2009ab). "he got game" theory: Optimal decision making and the nba. UCSD.

[Rao, 009b] Rao, J. M. (2009b). Seeing is believing: An explanation for belief in the hot hand. UCSD.

[Rao, 009c] Rao, J. M. (2009c). When the gambler's fallacy becomes the hot hand fallacy: An experiment with experts and novices. UCSD.

[Romer, 2007] Romer, D. (2007). It's fourth down and what does the bellman equation say? a dynamic programming analysis of football strategy. NBER Working Paper No. W0924.

[Rust, 1987] Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5):999–1033.

[Santos-Pinto and Sobel, 2005] Santos-Pinto, L. and Sobel, J. (2005). A model of Positive Self-image in Subjective Assessments. *American Economic Review*, 95(5):1386–1402.

[Schuster et al., 2005] Schuster, M., McGlynn, E., and Brook, R. (2005). How Good is the Quality of Health Care in the United States? *The Milbank Quarterly*, 83(4):843–895.

[Selten, 1998] Selten, R. (1998). Axiomatic Characterization of the Quadratic Scoring Rule. *Experimental Economics*, 1(1):43–62.

[Shachat, 2002] Shachat, J. (2002). Mixed strategy play and the minimax hypothesis. *Journal of Economic Theory*, 104(1):189–226.

[Simons et al., 1970] Simons, H., Berkowitz, N., and Moyer, R. (1970). Similarity, Credibility, and Attitude Change: A Review and a Theory. *Psychological Bulletin*, 73(1):1–16.

[Sirri and Tufano, 1998] Sirri, E. R. and Tufano, P. (1998). Costly search and mutual fund flows. *The Journal of Finance*, 53(5):1589–1622.

[Taylor and Brown, 1994] Taylor, S. and Brown, J. (1994). Positive Illusions and Well-being Revisited: Separating Fact from Fiction. *Psychological Bulletin*, 116:21–21.

[Terrell, 1994] Terrell, D. (1994). A test of the gambler's fallacy: Evidence from pari-mutuel games. *Journal of Risk and Uncertainty*, 8(3):309–17.

[Trivers and Willard, 1973] Trivers, R. and Willard, D. (1973). Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring. *Science*, 179(4068):90.

[Tversky and Kahneman, 1974] Tversky, A. and Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131.

[Tversky and Villone, 1989] Tversky, A. and Villone, R. (1989). The hot hand: Statistical reality or cognitive illusion. *Chance*, pages 36–41.

[Varian, 1989] Varian, H. (1989). Differences of opinion in financial markets. In *Financial Risk: Theory, Evidence and Implications: Proceedings of the 11th Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis*, pages 3–37.

[Walker and Wooders, 2001] Walker, M. and Wooders, J. (2001). Minimax play at wimbledon. *The American Economic Review*, 91(5):1521–1538.

[Wardrop, 1995] Wardrop, R. L. (1995). Simpson's paradox and the hot hand in basketball. *Cognitive Psychology*, 49(1).

[Weeks et al., 1998] Weeks, J., Cook, E., O'Day, S., Peterson, L., Wenger, N., Reding, D., Harrell, F., Kussin, P., Dawson, N., and Connors, Jr, A. (1998). Relationship Between Cancer Patients' Predictions of Prognosis and Their Treatment Preferences. *Journal of the American Medical Association*, 279(21):1709–1714.