

Evaluating Player Strategies in the Design of a Hot Hand Game

Paul Williams, Keith Nesbitt, Ami Eidels, Mark Washburn, David Cornforth

Abstract— The user’s strategy and their approach to decision-making are two important concerns when designing user-centric software. While decision-making and strategy are key factors in a wide range of business systems from stock market trading to medical diagnosis, in this paper we focus on the role these factors play in a serious computer game. Players may adopt individual strategies when playing a computer game. Furthermore, different approaches to playing the game may impact on the effectiveness of the core mechanics designed into the game play. In this paper we investigate player strategy in relation to two serious games designed for studying the ‘hot hand’. The ‘hot hand’ is an interesting psychological phenomenon originally studied in sports such as basketball. The study of ‘hot hand’ promises to shed further light on cognitive decision-making tasks applicable to domains beyond sport. The ‘hot hand’ suggests that players sometimes display above average performance, get on a hot streak, or develop ‘hot hands’. Although this is a widely held belief, analysis of data in a number of sports has produced mixed findings. While this lack of evidence may indicate belief in the hot hand is a cognitive fallacy, alternate views have suggested that the player’s strategy, confidence, and risk-taking may account for the difficulty of measuring the hot hand. Unfortunately, it is difficult to objectively measure and quantify the amount of risk taking in a sporting contest. Therefore to investigate this phenomenon more closely we developed novel, tailor-made computer games that allow rigorous empirical study of ‘hot hands’. The design of such games has some specific design requirements. The gameplay needs to allow players to perform a sequence of repeated challenges, where they either fail or succeed with about equal likelihood. Importantly the design also needs to allow players to choose a strategy entailing more or less risk in response to their current performance. In this paper we compare two hot hand game designs by collecting empirical data that captures player performance in terms of success and level of difficulty (as gauged by response time). We then use a variety of analytical and visualization techniques to study player strategies in these games. This allows us to detect a key design flaw the first game and validate the design of the second game for use in further studies of the hot hand phenomenon.

Index Terms—Evaluation, Games, Psychology, User-centered design.

Manuscript received May 31, 2013.

P. Williams is with the University of Newcastle, Newcastle, 2299, Australia.

K. Nesbitt is with the University of Newcastle, Newcastle, 2299, Australia. (e-mail: keith.nesbitt@newcastle.edu.au)

A. Eidels is with the University of Newcastle, Newcastle, 2299, Australia.

M. Washburn is with the University of Newcastle, Newcastle, 2299, Australia.

D. Cornforth is with the University of Newcastle, Newcastle, 2299, Australia.

I. INTRODUCTION

DECISION-MAKING, risk-taking and strategy are important dimensions to many key business tasks, including trading shares, buying and selling real estate, project management and medical diagnosis. This paper examines a particular facet of decision making related to sports called the ‘hot hand’. While the domain under study is sports-related the outcomes promise to be more generally applicable to software in more traditional business domains. This work also provides an interesting case study in the use of serious computer games to study decision-making. During the development of these games the interesting question of how unexpected user strategies might impact on outcomes is raised. Furthermore the outcomes highlight the importance of using empirical data to test user strategy when developing software.

Computer games often require players to exert significant perceptual and cognitive effort to be successful. This effort has been harnessed for tasks such as predicting the structure or proteins [1], labeling objects in images [2] and recognizing parts of images [3]. Computer games have also been widely spoken of as new multimedia platforms for general learning [4] and communicating about science [5].

There is also a significant potential for using computer games to assist with psychological research. Indeed a number of studies have used existing games such as Tetris and Madden to explore aspects of cognition [6, 7, 8]. Game engines have also been used to support studies in spatial cognition and social behavior [9, 10, 11, 12, 13].

In this paper we describe the development of two serious games to assist in the study of the psychological phenomenon known as the ‘hot hand’ [14]. To be useful in such a study these games need to meet particular design criteria in terms of player performance. In interface terms this performance is related to the efficiency and effectiveness of the user. As in typical usability studies we gathered empirical data under experimental conditions to test that our games meet our design criteria. Using this approach we found that the first game had an unintentional design flaw. This flaw made it less suitable for studying the hot-hand phenomenon. Therefore we developed a second game to address this problem. After following a similar empirical testing procedure the second game was found to meet our hot-hand requirements.

In the next section we discuss the hot-hand phenomenon and the particular design requirements for a game that allows the study of the hot hand. In the subsequent sections we

describe our first game design, called 'Aliens', the methods we used to test it, and the results of our usability analysis. We then describe a second game design, called 'Buckets', and provide an analysis of results from this study in a similar manner. In the final section of the paper we compare and contrast the results from the two game designs and discuss directions for future work.

A. Hot Hand

The term 'hot hand' describes the belief that the probability of a hit (success) following a hit should be greater than the probability of a hit following a miss (failure). In seminal research, it was found that 91% of basketball fans believed professional players had a better chance of making a shot after having hit their previous two or three shots than after having missed their previous two or three shots [15]. Professional basketball players also endorsed the belief; with each interviewed agreeing "it was important to pass the ball to a player who had made several shots in a row" (p. 302).

While intuitively these beliefs and predictions seem reasonable, no evidence for the hot hand was found in the field-goal shooting data of the 1980-81 Philadelphia 76ers. Likewise, further analysis of data from professional basketball [16], baseball [17] and golf [18, 19, 20] all failed to support the intuitive belief in the hot hand. This lack of empirical evidence led some theorists to suggest that the belief in the hot hand is a psychological fallacy [15, 21, 22]. That is, hot and cold streaks in performance are a myth that players and spectators endorse.

The most common explanation for the disparity between the popular belief that hot hand exists and actual data that shows no support for hot hand is that humans tend to misinterpret patterns in small runs of numbers [15]. That is, we tend to form patterns based on a cluster of a few events, such as a player scoring three shots in a row. We then use these patterns to help predict the outcome of the next event, even though there is insufficient information to make this prediction [23]. This is somewhat akin to the 'gamblers fallacy' that also arises from a belief in the law of small numbers [24], although for reasons we shall not discuss here the latter actually makes opposite predictions (people expect gamblers to fail after successful streaks).

However, the somewhat elusive hot-hand effect has been reported in the literature. Players have been reported to get on hot streaks in tasks such as horseshoe pitching [25], billiards [26] and ten-pin bowling [27]. Most recently, [28] found strong evidence for hot hand performance in volleyball. Although early hot hand findings (i.e., the lack of hot hand) seemed at odds with intuitive predictions, there now seems to be more to the hot hand picture than can simply be explained by a cognitive fallacy.

Under close examination, empirical studies of the hot hand seem to follow a qualitative pattern. On the one hand (no pun intended), in tasks where the difficulty of each shot is largely 'fixed' the hot hand seems common. This is true in tasks like horseshoe pitching and ten-pin bowling. Even in games like

volleyball the defensive side must remain on the opposite side of the net and cannot influence the striker greatly. On the other hand, in sports where the difficulty of each shot attempt is 'variable' there is no evidence in the data for hot or cold streaks. This is true in sports like basketball where the defense can interfere.

Hot hand may be a myth resulting from a cognitive fallacy, however, the pattern highlighted by grouping 'fixed' and 'variable' studies seems to support alternative interpretations. One such explanation was provided by Smith [25] who suggested shooters might systematically take more difficult shots in response to a run of hits. Under this scenario, a player does show an increase in performance during a hot streak - as they are performing a more difficult task at the same level of accuracy. This increase in performance would not be detected by traditional accuracy measures, but may be detected by teammates and spectators.

While this hypothetical *difficulty-account* receives tentative support by drawing a distinction between fixed and variable difficulty tasks (as the hot hand is more likely to appear in fixed-difficulty tasks, where players cannot engage in a more difficult shot), further support must be provided for two underlying assumptions. These assumptions are that (1) when task difficulty is considered, players' performance can be different to what is predicted or expected, and (2) that people sometimes take on more difficult tasks in response to success and easier tasks in response to failure.

The first assumption under investigation can be framed in terms of a difficulty-accuracy trade-off. To explain, consider that as a task becomes more difficult, people tend to perform the task with less accuracy. Is it possible however that people performance might differ from this intuitive difficulty-accuracy trade-off? More specifically even, can people maintain performance levels as a task becomes more difficult?

Psychological research suggests this is possible. In fact two prominent groups of cognitive theories account for such findings. Energetical theories [29, 30] suggest increases in task difficulty lead to an increase in arousal, which in turn increases the maximum level of mental effort available to a task. On the other hand, perceptual load theory [31] suggests high perceptual load (i.e., higher difficulty) leads to a decrease in distraction from other information, thus allowing greater focus on more difficult tasks. A large body of evidence supports this account in perception [32].

Importantly these findings are not restricted to the laboratory. For instance in a famous study on Munich taxis, half of a fleet of otherwise identical taxis were fitted with an anti-lock braking system (ABS). ABS brakes improve driver control under braking [33], and as a result make driving easier and safer. However, over a 12 month period in which distance travelled and driver ability were controlled, no difference was found in the number or severity of accidents for taxis with and without ABS brakes. Wilde [34] suggested this and other similar findings demonstrate that people are willing to accept a consistent level of risk. They will maintain this fixed risk level even when conditions vary; in the taxis study, for example, the level of driving errors (risk) had

remained constant despite safer driving conditions. Likewise, Wilde [34] argued that if tasks become more difficult people might become more careful. By managing risk in this way, it is plausible that people can maintain consistent accuracy across different levels of difficulty.

The second assumption that requires support involves people's reactions to success and failure. In essence - is there evidence to suggest that people may attempt more difficult tasks after successes, and less difficult tasks after failures?

Experimental support provides some evidence for this claim. Wilde, Gerszke, and Paulozza [35] asked participants to tap a series of red squares after their appearance on a computer screen. Responses closest to 1500ms were rewarded the highest points, however responses faster than 1500ms were penalized. In response to a run of point scoring trials, participants adopted successively more risk by making faster taps, however, following a penalty, subsequent taps were significantly slower and less risky. These results suggest people may take riskier options after successes, and less risky options after failures. It follows that performers may systematically adjust task difficulty in response to success and failure.

Attempts have also been made to assess this assumption outside of the laboratory. Rao [36] analyzed 60 LA Lakers basketball games in the 2007-08 season, and reported that while the majority of players attempted more difficult shots following a successful run, no tendency was found for players to attempt less difficult shots following an unsuccessful run. While Rao's results are of interest, the complexities of sports analysis must be considered. It is debatable whether any coding system can accurately assess the variety of contexts in which basketball shots are taken; particularly given individual players differ in shooting strengths and weaknesses.

We are thus faced with a dilemma. It seems we can support our two key assumptions, however more data needs to be gathered to investigate potential explanations of the hot hand. Unfortunately trying to gather more data from sporting games and contests is fraught with problems of subjectivity. How can one objectively assess the difficulty of a given shot over another in basketball? How can one accurately tell if a player is adopting an approach with more risk?

Our proposed solution to this problem is to design computer challenges of matched 'variable' and 'fixed' difficulty tasks that can be employed to test various hypotheses surrounding the hot hand. This presents challenges in designing tasks or game challenges that have particular usability characteristics. This paper focuses on the characteristics required in variable-difficulty hot-hand games.

A variable difficulty hot-hand game requires some careful design and testing if it is to be used to gain insight into how players respond to a run of success or failure. Namely, a hot hand game must provide a challenge with binary outcomes, that is, a challenge in which a player either succeeds or fails. The player must also be given clear feedback on each outcome, the same way a basketball player knows for sure whether he had hit or missed.

We intend to use the game to study a precise psychological

phenomenon related to hot and cold streaks in performance. Therefore, a further requirement for a hot hand game is that it allows measurement of players' strategy after runs of both successes and failures. If people fail most of the time, we won't record enough runs of success. If people succeed most of the time, we won't observe enough runs of failure. Thus, the core challenge needs to provide a probability of success, on average, somewhere in the range of 40-60%.

However the most significant requirement for a hot hand game is that it requires a finely tuned risk and reward structure [37]. The game must allow players to take risks and to be adequately rewarded for the risk. If, for example, one risk level provides substantially more reward than any other, players will learn this reward structure over time, and be unlikely to change strategy throughout play. We would thus like each risk level to be, for the average player, equally rewarding. In other words, regardless of the level of risk adopted, the player should have about the same chance of obtaining the best score. In the games described below this is managed by balancing speed in the task with accuracy. The faster a player responds the less likely they are to succeed. This is balanced by allowing more opportunities for the player to attempt the task when they respond faster. So even though at faster speeds players may make more errors they will receive more chances to succeed.

In this paper we outline the development and analysis of two 'variable' difficulty tasks. One for a game called Aliens, and the other for a game called Buckets. The tasks in these games are designed so that changes in player strategies can be accurately recorded as the game progresses. We compare the two game designs by collecting empirical data that captures player performance in terms of success and shot difficulty (response time). In terms of usability these measures equate to effectiveness and efficiency.

Having collected the data we then used a variety of analytical and visualization techniques to study player strategies in these games. This allowed us to detect a key design flaw in the Aliens game, which made the game less suitable for hot-hand investigation, leading to the design of the Buckets game. Testing of this game showed that we had successfully removed the flaw and the resultant game was suitable for further study of the hot hand.

II. EXPERIMENT 1: 'ALIENS'

A. *The Aliens Game*

The Aliens game is a simple first person shooter game developed in Flash and Actionscript (see Fig. 1). The players' goal is to shoot down as many alien spacecraft as possible within the overall time allowed. The game consists of a repeated challenge where a single alien spacecraft appears and moves across the game screen and the player's spacecraft is allowed a single shot to hit the alien. Entry and exit by each new alien (a trial) can therefore result in a hit or miss.

All trials are separated by a brief period where no alien is on the screen. New trials always begin unless time has run out. In the case where a trial is underway as time runs out, the trial

continues until completion but no result is recorded. For each new trial the player's spacecraft is fixed in a random position within an area ± 100 pixels from the screen center (see Fig. 2).



Fig. 1. Screenshot of the Aliens game in operation.

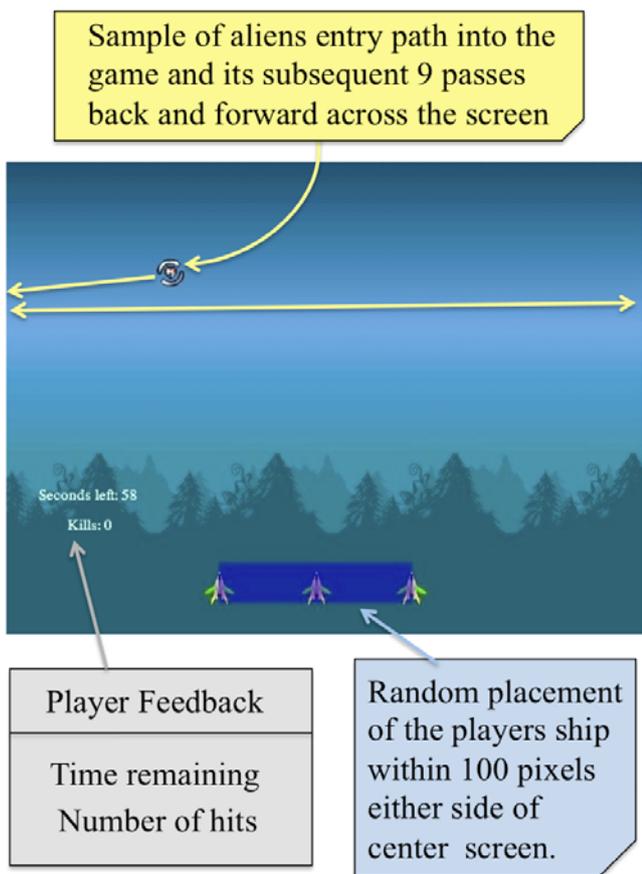


Fig. 2. The main mechanics of the Aliens game.

On each trial, an alien spacecraft (hereupon alien) enters the top of the game screen and moves either left or right in a downward arc until reaching a set height from the top of the screen (see Fig 2). The alien then travels from side to side at

this height passing over the player-shooter a maximum of nine times. The aim for a player is to time their shot so that a bullet from the player's spacecraft intercepts the alien on one of these nine passes. A shot is declared a miss once the bullet clears the maximum height of the alien without making contact. A shot is declared a hit if the bullet intercepts the alien (pixel contact). If a player fails to take a shot during the nine passes of the alien then it is considered a non-attempt.

The player is allowed only one shot per alien. If an alien is hit, it explodes onscreen. Each shot and hit is accompanied by appropriate auditory effect. The trial completes immediately if the player is successful with their shot. If a shot is missed, a penalty period ensues while the alien completes the nine passes and exits the screen.

Importantly, in each successive pass the alien spacecraft decelerates. An assumption is that the slower the alien moves the easier it becomes to target. Therefore, the longer a player waits to take her shot, the easier it becomes to hit the alien. Since a player is allowed only one shot per alien, the game incorporates an element of strategy - shooting on earlier passes allows more time for additional attempts at shooting aliens. However, earlier passes present more difficult shots, increasing the players' risk of failure.

For all complete trials the initial direction of the alien (left/right), the position of the player-shooter (± 100 pixels from the center), the difficulty of the attempted shot (pass number 1-9, where 1 indicates the highest shot difficulty), and the outcome of the shot attempt ($h = 1$; $m = -1$; non-attempt = 0) are recorded. The block and trial number are also recorded for each shot.

B. Methods

The experiment was run in a dimly lit room on IBM compatible computers using Windows XP and standard keyboards. Seventeen-inch CRT monitors were used for the experiment with screen resolution set to 1024 by 768 pixels. The experiment was coded in Actionscript 3.0 and run in Mozilla Firefox version 3.1 browsers for Windows. Participants wore Sennheiser headphones.

Twenty-nine participants from the University of Newcastle, Australia volunteered in response to recruitment posters. All participants had normal hearing and normal or corrected-to-normal vision. Each of the participants in the study was reimbursed \$AUS10 for taking part in the experiment.

Participants first played two three minute time periods (blocks) for training purposes. After these training blocks they played a further three blocks of trials each lasting 12 minutes. To progress between each block, participants had to press the spacebar. The spacebar was also used in the game to fire each bullet. Participants were asked to maintain a comfortable, self-selected distance from the screen throughout.

Both verbal and onscreen instructions outlined the goal and rules of the game for participants. Players were advised that the first two blocks should be treated as "practice", and that shooting down as many aliens as possible would require "both speed and accuracy".

A simple visual interface provided feedback on the player's

current performance and game status by registering the number of kills (hits) and the time remaining during each block (game) (see Fig. 2). At the completion of a block, participants also received summary feedback on the number of kills made for that block, and their grand total number of kills during the experiment. Participants were encouraged to use this feedback to monitor performance and set future goals.

C. Results and Discussion

A summary of results for the participants is shown in TABLE I. On average each participant in the study completed 407 trials. The average number of hits was 151 and the average number of misses was 256. However, there were large variations in player performance. For example, in terms of the number of trials completed there was a standard deviation of approximately 146. Indeed the maximum number of completed trials was 810 and minimum was 255.

To look for clusters of players who performed at different levels of expertise, or who used different strategies, we calculated each player's percentage success rate, their average response time (as a gauge for shot difficulty), and their total number of hits. We then used this data in a multi-dimensional scaling routine based on a Sammon projection [38]. The results of our Sammon mapping are shown in Figure 3. As can be seen in this figure, players 3, 6, 8, 9 and 21 appear as a unique cluster in terms of their performance.

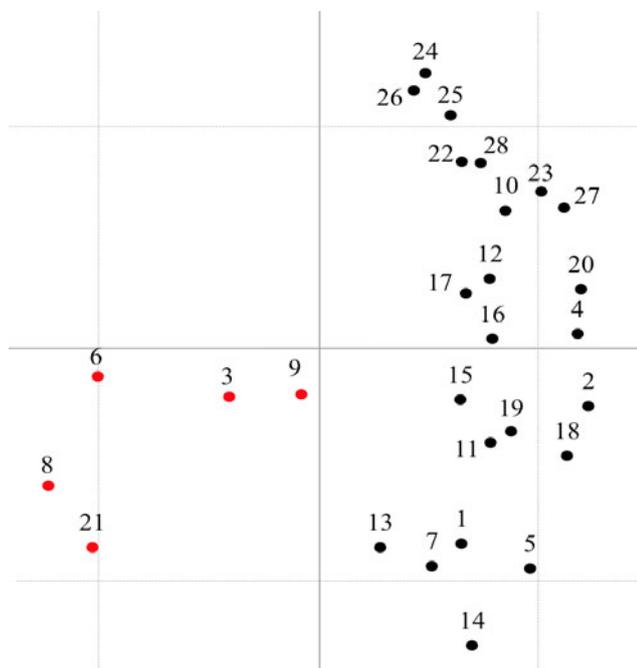


Fig. 3. Example Multi-dimensional scaling from the Sammon mapping indicating a distinct cluster of players (3,6,8,9,21)

While the non-linear projection associated with the Sammon mapping is difficult to correlate with the original variables it is extremely useful for the type of exploratory analysis we wanted to perform. Once we identified two possible player clusters we then used further interactive

visualization software to analyze the players in terms of response times, success rates and the total number of hits (see Fig. 4).

TABLE I. NUMBER TRIALS PER PLAYER, AVERAGE PLAYER HIT RATES AND RESPONSE TIMES IN THE ALIENS GAME. FIVE PLAYERS (3,6,8,9,21) COMPLETED WELL ABOVE THE AVERAGE NUMBER OF HITS (151).

Player	Number Trials	% Hits	Average Response Time (pass no)	Total Hits
1	415	29	4.7	121
2	274	31	7.4	86
3	457	45	4.3	204
4	255	36	8.1	93
5	369	25	5.1	93
6	532	49	3.5	261
7	564	25	4.8	141
8	720	42	2.8	300
9	510	40	5.5	203
10	299	44	9.1	133
11	434	30	6.6	131
12	353	40	8.4	142
13	582	28	4.6	162
14	617	19	4.2	120
15	421	34	6.8	143
16	368	37	7.7	135
17	361	40	8.0	146
18	395	27	7.2	105
19	416	30	6.9	126
20	303	36	9.0	109
21	802	36	2.7	286
22	276	50	9.1	138
23	255	45	9.4	115
24	279	58	9.8	161
25	283	53	9.4	150
26	291	56	9.4	163
27	258	43	9.6	110
28	298	49	9.2	145

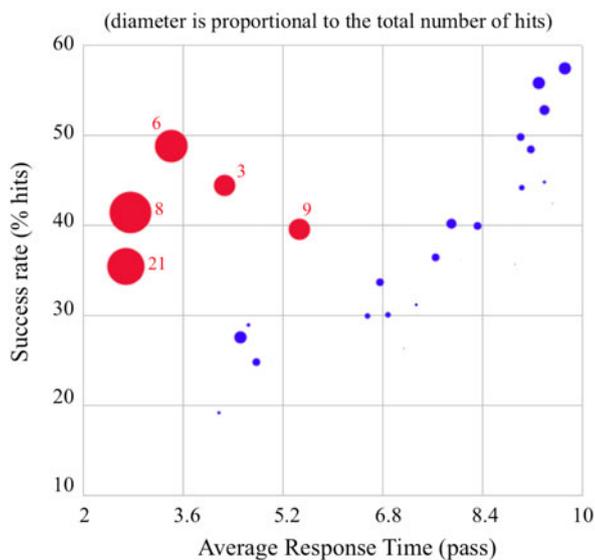


Fig. 4. Average response time versus the success rate for players in the Alien game. The diameter of points on the plot shows the relative number of hits during the game. The 5 players indicated are characterised by a high number of hits, low response time and unexpectedly high success rates.

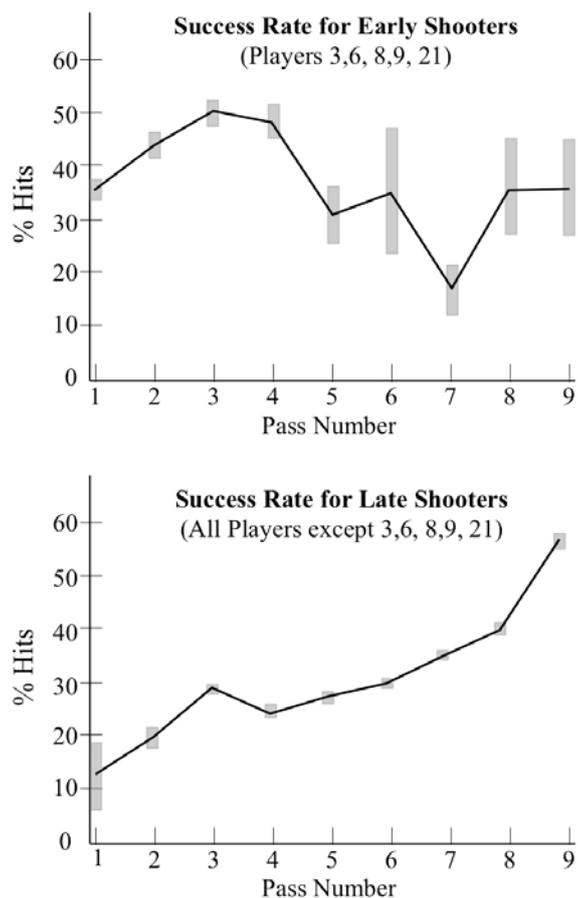


Fig. 5. Early versus late shooters in the Aliens game showing hit rate by pass for all shots taken by players 3,6,8,9,21 (top) with all other players (below). Notably, early shooters found shooting at early passes easier than late passes. Late shooters reflect expected difficulty, being successful at later passes.

To try and understand how more difficult, early shots could result in a higher probability of hits for some players we interviewed some of the identified group. They had discovered that on early passes, they could accurately time their shot by watching the approach of the alien to the edge of the screen. This provided in effect a low risk, high reward way to shoot early in the trial. The edge of the screen acted as a kind of ‘gun sight’. It seems that four other players also identified the same strategy. The effect of this strategy is shown in Fig. 5.

Unfortunately this unintentional flaw in our Alien game design made it unsuitable for testing the hot hand phenomenon. The risk and reward for players of a hot hand game need to be balanced so that higher risk behavior from the player incurs lower levels of reward. As a result of this problem we designed an alternative game based on a simple perceptual challenge. This second game was called Buckets and is described in the next section.

III. EXPERIMENT 2: ‘BUCKETS’

A. The Buckets Game

The Buckets game is based on a repeated perceptual challenge that requires players to decide which of four buckets is becoming darker (see Fig 6.). The goal of the game is to identify as many target buckets as possible in a fixed time period.

At the beginning of each trial, players view four buckets (rectangles). Each bucket is half filled with blue pixels (drops) that have been randomly positioned. This display is shown in Fig. 6. During a trial, the blue pixels are randomly repositioned 10 times per second, creating a dynamic effect within every bucket similar to visual static. Over the course of a trial, one bucket (the target) accumulates additional blue pixels at a constant rate. Players can attempt to select the target at any time. A correct target selection is declared a hit, while an incorrect detection is declared a miss. Players are provided with clear visual and auditory feedback signaling the outcome of each trial.

The response time of the player in the Buckets game is equivalent to the pass number measured in the Aliens game. The faster a player responds, the more difficult the task should become. Hence faster decisions allow more time for additional trials, however faster decisions are more risky and may be more likely to result in failure. This has been achieved by allowing more dark pixels (drops) to accumulate in the target bucket as the trial progresses. Drops accumulate at a constant rate in the target bucket, so as time progresses it becomes easier to distinguish from the three distracter buckets that do not accumulate more drops over time. In this way the Buckets Aliens games are analogous – a risk/reward strategy must be adopted in both games with the aim of finishing as many correct trials as possible within a fixed time period.

Despite these similarities, the games have an important difference. In the Buckets game a player has a 25% chance of simply guessing and still identifying the target correctly. Therefore, a player could attempt many trials and make many successes by simply guessing at the earliest possible time on

every trial. To counteract this strategy, incorrect decisions were followed by a brief penalty time period before the next trial began.

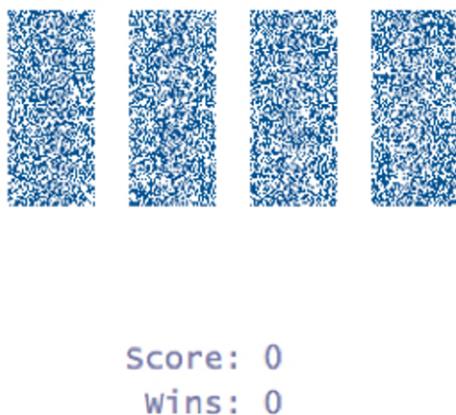


Fig. 6. The Buckets Screen showing the four buckets partially filled.

Once again the design of the game play emphasizes the need for both accuracy and speed in the players' responses. Waiting for the trial to be easy incurs a time cost, reducing the overall time remaining for subsequent attempts. Importantly, the game mechanics (i.e., rate of introduction of pixels, penalties, etc) were extensively piloted so that early attempts would provide roughly the same amount of correct decisions as later attempts over a long time period. A simple scoring mechanism keeps count of the number of wins or hits and provides feedback to the player (see Fig 6.).

Trials are separated by a brief period where no buckets are on the screen. New trials always begin unless time has run out. In the case where a trial is underway as time runs out, the trial continues until completion but no result is recorded. For all other trials the difficulty of the attempted shot (response time, where closer to 0 indicates the highest shot difficulty), and the outcome were recorded.

B. Methods

Twenty-four participants with normal or corrected to normal vision were recruited via posters placed at the University of Newcastle. In this game, each player was paid a set amount per correct response to help motivate them to make as many correct target selections as possible.

Before play, participants were shown two complete trials that did not require any response. This allowed them to view the total amount of change in the target over the course of each trial. All participants then played a 5-minutes long practice block, followed by four experimental blocks of 10 minutes each. Participants were encouraged to explore differing strategies during practice, and were only paid per correct response during the experimental blocks. Again the game goals were explained verbally and onscreen.

A complete trial uninterrupted by a player's response lasted 8000 ms (80 updates). Additional blue pixels were introduced at 1.875 pixels per update. Other game variables included a 300ms central fixation cross before each trial, 250ms pre- and

post-fixation blank screens, and feedback after attempts (500ms for correct and 2150ms for incorrect attempts; the difference of 1650ms being the penalty for incorrect decisions). Participants indicated which rectangle they believed was the target by pressing one of four spatially-corresponding keys ('a', 's', ';', or '''), with each success being worth 1 point. At the end of each block, participants were given feedback on the number of correct decisions made for that block and their grand total. They were encouraged to use this feedback to monitor their performance.

The experiment was again run on IBM compatible computers using Windows XP and standard keyboards. Screen resolution was set to 1024 by 768 pixels on 17" CTR monitors. The experiment was coded in Actionscript 3.0 and run in Mozilla Firefox version 3.1 browsers for Windows. Participants wore Sennheiser headphones.

C. Results and Discussion

On average each participant completed approximately 370 trials with an average of 210 hits and 160 misses. There was a standard deviation in the number of trials of approximately 30. We note that the variation between players in terms of completed trials was much lower than the variable performance seen in the Aliens experiment. The key results for each player in the Buckets experiment are shown in TABLE II.

TABLE II.

AVERAGE PLAYER HIT RATES AND RESPONSE TIMES IN THE BUCKETS GAME.

Player	% Hits	Average Response Time (ms)	Total Hits
1	64	5.3	222
2	54	3.9	226
3	73	5.4	255
4	63	4.9	234
5	32	2.3	176
6	54	5.2	189
7	68	4.8	256
8	51	5.5	169
9	55	4.4	215
10	52	4.0	213
11	60	4.9	220
12	37	3.9	150
13	66	5.2	234
14	64	5.8	211
15	62	5.5	213
16	60	5.5	203
17	71	5.6	241
18	59	4.9	215
19	51	6.5	151
20	53	4.7	196
21	72	5.9	238
22	69	5.1	251
23	49	4.6	183
24	46	4.1	180

Once again we looked for clusters of players using the normalized results for each player's percentage success rate, their average response time and their total number of hits. We followed the same procedure as in the Aliens game and used these data in a multi-dimensional scaling routine based on a Sammon projection [38]. The results of our Sammon mapping for Experiment 2 are shown in Fig. 7. As can be seen in this figure, there appeared to be only one main cluster although players 5 and 19 appeared to be outliers. Table 2 highlights the results from these same two players.

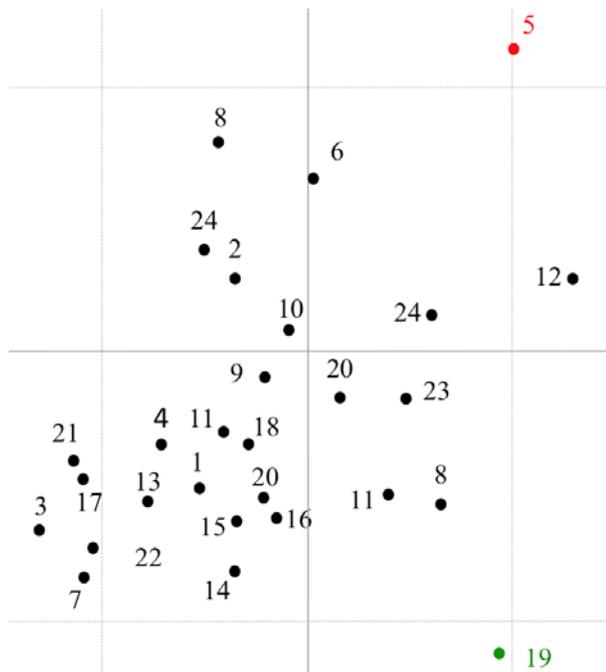


Fig. 7. Multi-dimensional scaling results from the Buckets game. Identifying two outliers (players 5, 19).

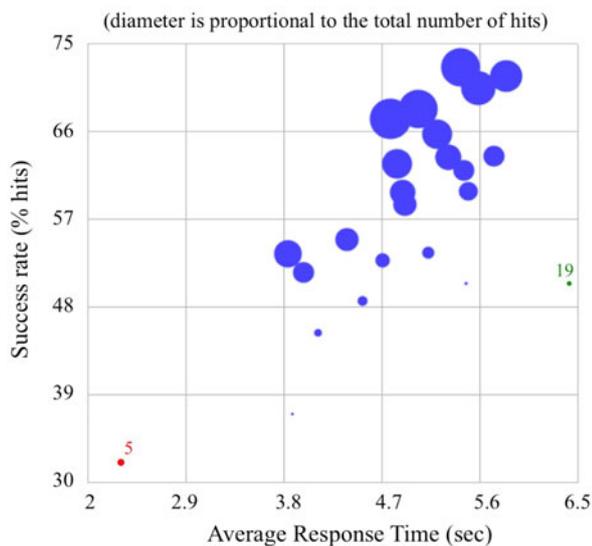


Fig. 8. Visualising player strategy in buckets game. Note that in comparison to Fig. 4 higher success rates are associated with slower response times.

Once more we followed the same procedure used with the Aliens game and employed interactive software to visualize the players in terms of their response times, success rates and total number of hits. Fig. 8 helps to highlight the two identified outliers. Player 5 seems to have shot very early, however as expected, she or he also had a low hit rate. That is this player took high risk but in doing so registered a low number of hits. Player 19 shot late but had a relatively low success rate. This may indicate poor aptitude to the task. The Pearson coefficient of correlation for average response time and percentage of hits using all players was 0.7. This satisfactory relationship between response time and success rates supports the use of the Buckets game in our hot hand study.

IV. GENERAL DISCUSSION

In this paper we have described the development and testing of two games, the 'Aliens' and 'Buckets' games. These games were specifically designed to study the Hot Hand phenomenon, which has been extensively studied in psychological research. These games offer for the first time a well-controlled testing environment for a phenomenon that was measured, up-till-now, off the laboratory and was therefore sensitive to a number of contextual variables (but see [39] for preliminary investigation in that direction). Analysis of the first game revealed certain biases in players' strategies that deemed it less appropriate for testing the Hot Hand. Therefore, a second experiment was developed with an eye on these biases. Indeed similar analysis on the results of the second game revealed it is robust to changes in players' strategies, and can therefore be used in the psychological arena to test the mechanisms that underlie the belief (and potential existence) of the Hot Hand.

The term Hot Hand marks the common belief, in basketball and other sports, that the probability of making a shot given that the player had just made the previous shot (i.e., the probability of a hit given a hit) is greater than the probability of making a shot given a miss on the previous shot-attempt. While strong belief in the hot hand is well documented, empirical evidence for hot hand is rather sparse. In their seminal study, Gilloovich et al. [15] showed that even though both spectators and players strongly believed in the hot hand, professional basketball players were not more likely to make a shot if it was preceded by a successful attempt. Similarly, studies in other sports [e.g., 16, 17, 18, 19, 20] all failed to provide empirical support for the hot hand. However, we recount that many of these studies focused on field sports, where experimental control is minimal if not impossible. Studying the hot hand with specialized computer games, as we did here, allowed much better control of critical experimental factors. It was this intersection of performance in a task, and the difficulty of the task at hand that formed our departure point for the current study.

We developed two computer games that allow measuring both the performance of the players, and the difficulty level of each and every shot attempt. Both games featured challenges with binary outcomes, where players could either succeed or

fail on each trial. This type of binary challenge is essential for testing the hot hand.

Another important design feature for a hot hand game is that players have an average success rate of between 40-60% for the challenges. This should allow for both hot and cold streaks to be distinguished in the data. The Aliens game had an average success rate of 39% (std dev 9.9). The Buckets game had an average success rate within this range, of 57.5% (std dev 10.4). While these success rates are at the limit of what we would like they are both considered acceptable.

The most important criterion for our hot hand games is that players are rewarded appropriately for both efficiency and effectiveness in the repeated tasks they undertake. We tried to design both games so that there was a balance between risk and reward. We encouraged players to take risks (respond early) by rewarding them with more shot attempts. In setting up this reward structure we intended that higher risk would equate to lower success rates in the task. However, after collecting empirical data for the first, Aliens game we uncovered a serious design flaw. Some players had uncovered a 'cheat' in the game and were able to achieve high success rates when responding early in the game. This effectively made the game unsuitable for studying the hot hand. The alternative game called Buckets was developed and tested in the same manner. It was found to meet the requirement that fast response times relate to low success rates, thus making it acceptable for further study of the hot hand.

We plotted hit-rate data from both the Aliens and Buckets games as a function of difficulty (Figures 4 and 8, respectively) and furthermore visualized these data using Sammon projection (Figures 3 and 7) to identify clusters of players with similar and dissimilar strategies. The qualitative patterns in Figure 4 and 8 differ in a meaningful way; players in Figure 8 (Buckets) are roughly aligned along the main diagonal, suggesting that hit rate increased for players that were willing to wait longer, on average, before making a decision. Figure 4 (Aliens), in contrast, reveals some players that have responded very quickly yet were able to maintain a high level of performance. We referred to this sub-group of players earlier and suggested they have identified a 'cheat' in the game. We concluded that players in the first game may be divided to two groups based on their response strategies, whereas such division is unlikely to have happened in the second game.

Yet, the fact the players presumably used a single response strategy in one domain does not imply they may not differ in other aspects. In the remaining of the discussion we highlight interesting differences in players' performance and strategy that had been revealed by our analyses.

First, players differed in their competence level on both games. Figure 8 shows performance in the Buckets game, measured by % hit, as a function of difficulty (gauged by average response time). For a given level of task-difficulty, such as responses that were executed at around 4.7 seconds, on average, one player had a success rate of 53% while another had a success rate of 68%, with yet other players within this range. Clearly, for the same level of task difficulty

different players could perform rather differently (by as much as $[68-53] / 53 = 28\%$, in this example). Differences in player performance are not unexpected in games. Even as early as 1979 Atari recognized this and designed games such as Adventure [40] for the Atari 2600 to provide different difficulty levels.. More recently a number of games such as Max Payne [41] and Left 4 Dead [42] have incorporated techniques known as "challenge functions" [43] to dynamically adapt the difficulty of game play based on the current player performance.

Players also differed in the risk they were willing to take. Some players were willing to commit to a decision relatively quickly, responding by as early as 3.8sec, on average, while others had waited longer, some of them as long as 5.8sec (see Fig. 8 again). While fast responses clearly impacted performance by way of pushing hit rate down, these 'fast-to-respond' players seemed to have been willing to accept the risk associated with fast responses. The finding that the overall level of risk accepted by players showed large individual differences is commensurate with psychological research surrounding impulsivity and risk-taking [44, 45]. Indeed, a future avenue for research will be to critically assess the relationship between these psychological constructs and players' behaviour in our hot hand games. Of course the combination of risk-taking and difficulty is also an important consideration in the design of games. Indeed some attempts have already been made to dynamically adapt the game play difficulty by accounting for both player performance and their risk profile [46].

Finally, players may differ in the way they explore the game's environment. Some players may explore pay-offs across a range of difficulty levels, to test how to maximize gains, while others may settle on a given level of risk in an attempt to exploit known rewards. Hills, Todd, and Goldstone [47], as well as others, studied the trade-off between exploitation and exploration in mental strategies. We have addressed this issue elsewhere, in the more specific context of hot hand games [39]. In the current games, an exploration strategy may have allowed some players in the Aliens game to identify a 'cheat'. These players may have tried to respond across a range of latencies and discovered, either by chance or via systematic exploration, that early shots reward them with a high hit-rate, while also conserving time for additional shots. Critically, at least from the perspective of the hot-hand research, no such behavior was similarly rewarded in the latter, Buckets game.

V. CONCLUSION

To conclude, we developed and tested two games that allow assessing both performance and shot-difficulty in a hot hand challenge. If we assume there is variable difficulty in some sporting tasks then measures of sport performance, such as basketball shooting percentages, can sometimes be misleading. The novel contribution of the proposed games in that they provide a controlled testing environment, one that allows to accurately measure both performance outcomes (shooting percentage) as well as the difficulty of each shot.

Thus, we expect it to become a useful tool in the systematic exploration of the hot hand phenomenon. In this paper we focused not only on the evaluation of players' performance level, but also the evaluation of players' strategies, particularly in terms of risk taking. Players could have saved time by taking early shots with higher difficulty, or obtain higher accuracy rate on the expense of time if they were to wait until the trial became easier. Our analyses revealed individual differences across players in game-competence, risk taking, and possibly exploration-exploitation strategies. However, based on cluster analysis, the structure of the Buckets game makes it robust to these differences and therefore adequate as a platform for studying the elusive hot hand phenomenon.

The value of this work extends beyond the understanding of how players make decision in games and sporting contests. Many traditional business applications also rely on users making decisions, taking risks and adopting strategies. Consider applications of intra-day trading where market traders make many rapid decisions about how to trade stocks. For example, do stock traders develop 'hot hands', perhaps taking greater risks after a successful string of trades? More generally, what role does user strategy play on the efficiency or effectiveness of software designed to support business tasks? Is there an opportunity to improve the design of business software by gathering more empirical data and looking at user patterns? These are a few examples of open questions that form part of our larger study beyond sporting contests and computer games.

ACKNOWLEDGMENT

We would like to thank David Elliot from the Newcastle Cognition Lab for his invaluable work in coding the various versions of both the Aliens and Buckets games.

REFERENCES

- [1] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović, and Foldit players, "Predicting protein structures with a multiplayer online game", *Nature*. Vol. 466 Issue 7307, p 756–760, 2010.
- [2] L. von Ahn, and L. Dabbish, "Labeling images with a computer game", in CHI '04: Proc. 2004 Conf. Human Factors Comput. Syst., pp. 319–326, 2004.
- [3] L. von Ahn, R. Liu, and M. Blum, "Peekaboom: a game for locating objects in images", in CHI '06: Proc. SIGCHI Conf. Human Factors Comp. Syst., pp. 55–64, 2006.
- [4] A. Derryberry, "Serious games: Online games for learning". Retrieved 10 August 2012 from Adobe Resources website: http://www.adobe.com/resources/elearning/pdfs/serious_games_wp.pdf
- [5] A. Krotoski, "Serious fun with computer games", *Nature*. vol. 466 Issue 7307, p. 695, 2010.
- [6] J. T. Hansberger, C. D. Schunn, and R. W. Holt, "Strategy variability: How too much of a good thing can hurt performance", *Memory & Cognition*, vol. 34, pp. 1652–1666, 2006.
- [7] D. Kirsh, and P. Maglio, "On distinguishing epistemic from pragmatic action", *Cognitive Science*, vol. 18, pp. 513–549, 1994.
- [8] P. P. Maglio, M. J. Wenger, and A. M. Copeland, "Evidence for the role of self-priming in epistemic action: Expertise and the effective use of memory", *Acta Psychologica*, vol. 127, pp. 72–88, 2008.
- [9] T. P. Alloway, M. Corley, and M. Ramscar, "Seeing ahead: Experience and language in spatial perspective", *Memory & Cognition*, vol. 34, pp. 380–386, 2006.
- [10] J. Drury, C. Cocking, S. Reicher, A. Burton, D. Schofield, A. Hardwick, D. Graham, and P. Langston, "Cooperation versus competition in a mass emergency evacuation: A new laboratory simulation and a new theoretical model", *Behavior Research Methods*, vol. 41, pp. 957–970, 2009.
- [11] A. Frey, J. Hartig, A. Ketzler, A. Zinkernagel, and H. Moosbrugger, "The use of virtual environments based on a modification of the computer game Quake III Arena in psychological experimenting", *Computers in Human Behavior*, vol. 23, pp. 2026–2039, 2007.
- [12] G. Gunzelmann, and J. R. Anderson, "Location matters: Why target location impacts performance in orientation tasks", *Memory & Cognition*, vol. 34, pp. 41–59, 2006.
- [13] G. A. Radvansky, and D. E. Copeland, "Walking through doorways causes forgetting: Situation models and experienced space", *Memory & Cognition*, vol. 34, pp. 1150–1156, 2006.
- [14] M. Bar-Eli, S. Avugos, and R. Raab, (2006). Twenty years of "hot hand" research: Review and critique", *Psychology of Sport and Exercise*, vol. 7, pp. 525–553, 2006.
- [15] T. Gilovitch, R. Vallone, and A. Tversky, "The hot hand in basketball: On the misperception of random sequences", *Cognitive Psychology*, vol. 17, pp. 295–314, 1985.
- [16] P. D. Larkey, R. A. Smith, and J. B. Kadane, "It's okay to believe in the "hot hand", *Chance*, vol. 2, pp. 22–30, 1989.
- [17] S. C. Albright, "A statistical analysis of hitting streaks in baseball." *Journal of the American Statistical Association*, vol. 88(424), pp. 1175–1183, 1993.
- [18] R. D. Clarke, "Streakiness among professional golfers: Fact or fiction?" *International Journal of Sports Psychology*, vol. 34, pp. 63–79, 2003.
- [19] R. D. Clarke, "An analysis of streaky performance on the LPGA tour. Perceptual and Motor Skills", vol. 97, pp. 365–370, 2003.
- [20] R. D. Clarke, "Examination of hole-to-hole streakiness on the PGA tour". *Perceptual and Motor Skills*, vol. 100, pp. 806–814, 2005.
- [21] R. L. Wardrop, "Simpson's Paradox and the Hot Hand in Basketball." *The American Statistician*, vol. 49(1), pp. 24–28, 1995.
- [22] A. Wilke, and H. C. Barrett, "The hot hand phenomenon as a cognitive adaptation to clumped resources." *Evolution and Human Behavior*, vol. 30(3), pp. 161–169, 2009.
- [23] A. Tversky and D. Kahneman, "Judgement under uncertainty: Heuristics and biases". *Science*, vol. 185(4157), pp. 1124–1131, 1974.
- [24] A. Tversky and D. Kahneman "Belief in the law of small numbers". *Psychological Bulletin*, vol. 76 (2), pp. 105–110, 1971.
- [25] G. Smith, G. Horseshoe pitchers' hot hands. *Psychonomic Bulletin and Review*, vol. 10 (3), 2003.
- [26] R. M. Adams, "Momentum in the performance of professional tournament pocket billiards players." *International Journal of Sport Psychology*, vol. 26, pp. 580–587, 1996.
- [27] R. Dorsey-Palmateer and G. Smith, "Bowlers' hot hands." *The American Statistician*, vol. 58(1), pp. 38–45. 2004.
- [28] M. Raab, B. Gula, and G. Gigerenzer, "The Hot Hand Exists in Volleyball and Is Used for Allocation Decisions." *Journal of Experimental Psychology: Applied*, vol. 18(1), pp. 81–94, 2012.
- [29] G. R. J. Hockey, "Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework." *Biological Psychology*, vol 45, pp. 73–93. 1997.
- [30] D. Kahneman, *Attention and Effort*. New Jersey: Prentice-Hall, Inc 1973.
- [31] N. Lavie, "Perceptual load as a necessary condition for selective attention." *Journal of Experimental Psychology: Human Perception and Performance*, vol 21(3), pp. 451–468. 1995.
- [32] N. Lavie, "Distracted and confused?: Selective attention under load." *TRENDS in Cognitive Sciences*, vol 9(2), pp. 74–82. 2005.
- [33] C. M. Farmer, A. K. Lund, R. E. Trempe, and E. R. Braver, (1997). "Fatal crashes of passenger vehicles before and after adding antilock braking systems." *Accident Analysis and Prevention*, vol 29, pp. 745–757. 1997.
- [34] G. J. S. Wilde, *Target Risk* (1st ed.). Toronto, Ontario, Canada: PDE Publications. 1994.
- [35] G. J. S. Wilde, D. Gerszke, and L. Paulozza, "Risk optimization training and transfer." *Transportation Research Part F: Traffic Psychology and Behaviour*, vol 1(1), 77–93. 1998.
- [36] J. M. Rao, "Experts perception of autocorrelation: The hot hand fallacy among professional basketball players." Unpublished manuscript,

Department of Economics, University of California, San Diego, United States. 2009.

- [37] J. W. Sammon "A nonlinear mapping for data structure analysis." IEEE Transactions on Computers vol. 18, pp. 401–409, 1969.
- [38] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
- [39] P. Williams, K. Nesbitt, A. Eidels, and D. Elliot, "Balancing Risk and Reward to Develop an Optimal Hot-Hand Game". *Game Studies*, vol. 11(1). ISSN:1604-7982, 2011.
- [40] Atari Inc. Adventure. [Atari 2600], USA: Atari Inc. 1979.
- [41] Valve Corporation. Left 4 Dead. [Windows], USA:Valve Corporation. 2008.
- [42] Remedy Entertainment. Max Payne. [Windows], USA:Gathering Of Developers. 2001.
- [43] P Demasi, and A Cruz, "Online Coevolution for Action Games." Proceedings of The 3rd International Conference on Intelligent Games And Simulation. London. pp. 113–120. 2002.
- [44] M. Zuckerman, *Sensation seeking: Beyond the optimal level of arousal*. Hilldale, New Jersey: Erlbaum. 1979.
- [45] M. Zuckerman, M. (2007). "The sensation seeking scale V (SSS-V): Still reliable and valid." *Personality and Individual Differences*, vol 43(5), pp. 1303-1305. 2007.
- [46] G. Hawkins, K. Nesbitt, and S. Brown, "Dynamic Difficulty Balancing for Cautious Players and Risk Takers." *International Journal of Computer Games Technology*, vol. 2012, Article ID 625476, 10 pages, 2012. doi:10.1155/2012/625476.
- [47] T. T. Hills, P. M. Todd, and R. L. Goldstone, "Search in external and internal spaces." *Psychological Science*, vol 19, pp.676-682. 2008.

Paul Williams is undertaking a PhD in Cognitive Psychology at the University of Newcastle, under the supervision of Dr. Ami Eidels. He is interested in developing online gaming platforms suitable for the investigation of how people change their behaviour in response to successes and failures. He is currently focused on refining a novel paradigm to study the behavioral phenomenon's known as the "hot hand" and "post-error slowing".

Paul.Williams@newcastle.edu.au

Dr Keith Nesbitt is a Senior Lecturer at the School of Design, Communication and IT, University of Newcastle. His research focuses on issues of perception and cognition applied to the areas of visualization, multi-sensory displays, virtual environments, computer games and conceptual modeling. You can access his website here: <http://www.knesbitt.com>

Keith.Nesbitt@newcastle.edu.au

Dr. Ami Eidels is a Lecturer in Cognitive Psychology, at the School of Psychology, University of Newcastle. He is also a principle investigator in the Newcastle Cognition Lab. His research focuses on visual perception and attention, combined with computational and mathematical modeling. You can access his lab's website here: <http://newcl.org/eidels>

Ami.Eidels@newcastle.edu.au

Mark Washburn is a former honours student in the Newcastle Cognition Lab, School of Psychology, University of Newcastle. Under the supervision of Dr. Ami Eidels he helped to design, develop, and test the Buckets game.

Mark.Washburn@uon.edu.au

Dr David Cornforth is with the School of Design, Communication and IT, University of Newcastle. His research interests are in health information systems, pattern recognition, artificial intelligence, multiagent simulation, and optimization. He is convener of the Applied Informatics Research Group, University of Newcastle.

David.Cornforth@newcastle.edu.au

This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.