

# Post-error recklessness and the hot hand

Paul Williams\*    Andrew Heathcote\*<sup>†</sup>    Keith Nesbitt\*    Ami Eidels\*

## Abstract

Although post-error slowing and the “hot hand” (streaks of good performance) are both types of sequential dependencies arising from the differential influence of success and failure, they have not previously been studied together. We bring together these two streams of research in a task where difficulty can be controlled by participants delaying their decisions, and where responses required a degree of deliberation, and so are relatively slow. We compared performance of unpaid participants against paid participants who were rewarded differentially, with higher reward for better performance. In contrast to most previous results, we found no post-error slowing for paid or unpaid participants. For the unpaid group, we found post-error speeding and a hot hand, even though the hot hand is typically considered a fallacy. Our results suggest that the effect of success and failure on subsequent performance may differ substantially with task characteristics and demands. We also found payment affected post-error performance; financially rewarding successful performance led to a more cautious approach following errors, whereas unrewarded performance led to recklessness following errors.

Keywords: post-error slowing, hot hand, cognitive control, financial incentives

## 1 Introduction

The effects of recent outcomes on future performance have been the subject of considerable interest, mainly in two largely non-overlapping literatures about *post-error slowing* and the *hot hand*. Post-error slowing describes systematic increases in response time (RT) following an error in rapid choice tasks (Laming, 1968; Rabbitt, 1966a). The hot hand originated in sports, and describes an increase in the probability of success after previous success. The hot hand is often considered a fallacy as, despite the strong beliefs of spectators and players, the effect is not often empirically observed in professional sports (Gilovich, Vallone & Tversky, 1985; see also Avugos, Köppen, Czienskowski, Raab & Bar-Eli, 2013). Although the two phenomena are framed in terms of failure (post-error slowing) and success (the hot hand), both are measured by a difference between post-error and post-correct performance. From a measurement perspective, the key difference has been the primary dependent variable — RT for post-error slowing, and the probability of success for the hot hand. Recently, however, post-error slowing research has placed increased importance on the effect of errors on subsequent accuracy (e.g., Danielmeier & Ullsperger, 2011; Notebaert et al., 2009; Schroder & Moser, 2014). Hot hand research has also increasingly examined

whether sports players attempt more difficult (e.g., quicker) shots following success, which may obscure improved performance if it is measured solely by accuracy (Bocskocsky, Ezekowitz & Stein, 2014; Rao, 2009). Thus, research on the hot hand and post-error slowing taps related questions.

Empirically, there are distinct similarities between recent hot hand findings and well-established regularities found in post-error slowing research. Rao (2009) used video analysis and found that basketball players attempted more difficult shots following a successful run. More recently, Bocskocsky, Ezekowitz and Stein (2014) employed enhanced tracking technology and found players on a “hot run” take more shots of higher difficulty, and perform at above expected performance levels if shot difficulty is taken into account. Although it is debateable whether the difficulty of complex actions such as basketball shots can be precisely quantified, the increased difficulty of basketball shots following success resembles performance in rapid-decision tasks where gradual speeding (analogous to more difficult shots) is observed over runs of correct responses that precede an error (Dudschig & Jentzsch, 2009; Laming, 1968; see Luce, 1986, for a review).

From a theoretical perspective, post-error slowing (Laming, 1968, 1979; Rabbitt, 1966a, 1966b, 1969; Rabbitt & Rodgers, 1977) was initially considered the result of an increase in caution following errors. The *caution explanation* also suggested that following success less caution is exercised, and response times get faster (Dudschig & Jentzsch, 2009; Jentzsch & Dudschig, 2009; Laming, 1968). Formal models of decision-making response-time (Dutilh et al., 2012a) and cognitive control (e.g., Botvinick, Braver, Barch, Carter & Cohen, 2001) have since established that increased

---

ARC-Discovery Project grants to AH and AE, ARC Professorial Fellowship to AH Keats Endowment Fund grant to PW and AE.

Copyright: © 2016. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

\*The University of Newcastle School of Psychology, University of Newcastle, Callaghan, NSW, 2308, Australia. Email: paul.williams@newcastle.edu.au.

<sup>†</sup>The University of Tasmania

response times following errors can often be causally linked with a higher response criterion following instances of high response conflict, including errors.<sup>1</sup> The caution explanation aligns with the hot hand framework of Bocskocsky, Ezekowitz and Stein (2014), who noted that basketball players might be less cautious following successes. Hence, basketball players and experimental participants alike potentially employ less caution following success and more caution following errors. The level of caution adopted following success relative to failure is, therefore, central to both domains.

Despite these similarities, the hot hand and post-error slowing have typically been studied over greatly differing time scales and across very different environments. Post-error slowing research has been narrowly focused on simple and rapid choice tasks with high levels of experimental control (Yeung & Summerfield, 2012). In contrast, hot-hand research has mainly focused on uncontrolled sporting tasks (e.g., shooting a basketball) that unfold over longer (and often irregular) time scales. These narrow foci leave open questions in each field regarding the generalizability of findings. For example, Yeung and Summerfield (2012) questioned the degree to which the current body of post-error literature might scale up to explain decisions that are goal driven and temporally extended. Similarly, Bocskocsky, Ezekowitz and Stein (2014) found empirical evidence in support of theoretical speculation that basketball shooters attempt more difficult shots following success — yet the possibility that this finding generalizes as a behavioural regularity remains untested. In sum, each domain has had a narrow focus, and these foci are so widely separated that it is unclear whether post-error experimental findings might shed light on goal-driven behaviours in more complex environments (such as sporting performance), and vice versa.

To better assess the similarities and differences between post-error slowing and the hot hand, data are required that connect the two domains. Here we collected such data using the *Buckets game*, a computerized task, created by Williams, Nesbitt, Eidels, Washburn and Cornforth (2013), that utilises an intermediate time scale connecting post-error slowing and hot hand research. Participants were presented, on each trial, with four rectangular “buckets”, each half-filled with randomly positioned pixels. Over time, one of the buckets (target) accrues more pixels and gradually become fuller, while the other buckets (distractors) remain half filled. The task of identifying the target bucket, therefore, becomes easier as the trial progresses. The defining features of this game are that it presents temporally extended decisions (trials lasted up to 8 seconds), and that players

can elect to respond more quickly with less chance of being correct, or more slowly with a higher chance of being correct. That is, players self-selected the level of difficulty they were willing to assume for each attempt. The goal of the game is to maximise the number of correct decisions in a fixed time period. Hence, responding quickly offered the benefit of more attempts overall, but at the risk of lower accuracy. Williams et al. (2013) described in detail how the game’s timing and incentive system were tuned. Players were explicitly informed that they control the difficulty of each attempt and that they can trade-off between difficulty and speed to maximize their overall performance. Because the speed-accuracy trade-off was explicit, the task slow-paced, and each individual attempt was embedded within the overarching global context of maximising correct decisions in a limited space of time, the task lent itself to deliberative post-error adjustments. In the language of Yeung and Summerfield (2012), the task encouraged meta-cognitive judgments.

With respect to post-error slowing, the Buckets game allows us to assess post-error adjustments in a relatively simple but goal-driven task that unfolds over up to 8 seconds. With respect to the hot hand, the Buckets game allows expansion of the recent work of Bocskocsky, Ezekowitz, and Stein (2014), who found professional basketball shooters attempted more risky or difficult shots after previous successes. We can assess in a controlled environment whether this finding reflects a systematic behavioural trend. Furthermore, if players systematically adopt more or less risk following success or failure, we can assess how this affects detection of the hot hand.

An important consideration in using the Buckets game is that participants are motivated to achieve its goals. Psychologists and economists hotly debate the benefits of financial incentives and how such incentives influence intrinsic and extrinsic motivation (e.g., Read, 2005; Camerer & Hogarth, 1999). Less controversial is the empirical finding that financial incentives do alter performance systematically in cognitive tasks (Botvinick & Braver, 2015; Camerer & Hogarth, 1999). For mundane laboratory tasks, financial incentives improve motivation and performance (Cameron, Banko & Pierce, 2001; Camerer & Hogarth, 1999; Kounieher, Charron & Koechlin, 2009; Padmala & Pessoa, 2011). Further, monetary rewards seem to facilitate performance to a greater extent when incentives are contingent upon the level of performance (Bonner, Hastie, Sprinkle & Young, 2000). Botvinick and Braver (2015) described improvements in cognitive task performance due to financial incentives as a fundamental phenomena that links motivation to cognitive control. That is, financial incentives increase the level of cognitive control available for a task, which in turn improves performance.

Botvinick and Braver (2015) note that fluctuations in cog-

<sup>1</sup>Note in other instances, increased response times following errors have been linked to multiple causes (Dutilh, Forstmann, Vandekerckhove & Wagenmakers, 2013), the need to re-orient to the task following errors (Notebaert et al., 2009), or an increase of inhibition (Ridderinkhof, 2002).

nitive control linked to motivation are observed not only in overall performance but also at short, trial-by-trial, time scales. Indeed, post-error adjustments are typically considered a fundamental aspect of cognitive control (Botvinick, Braver, Barch, Carter & Cohen, 2001; Gehring & Fencsik, 2001; Ridderinkhof, Van Den Wildenberg, Wijnen & Burle, 2004). It is not surprising, then, that, like overall performance, post-error adjustments have been empirically linked to financial incentives and motivation. Sturmer, Nigbur, Schact and Sommer (2011) found that performance contingent on incentives led to an increase in post-error slowing, which is commensurate with findings of increased post-error slowing when financial rewards were tied to more accurate performance (Ullsperger & Szymanowski, 2004).

Motivation has also been of interest in the hot hand domain. For example, null results from experimental investigations (e.g., Gilden & Wilson, 1995) have been criticised because they were not collected from highly motivated participants typical of professional sporting settings (Smith, 2003). Thus, motivation and its effects on control and performance are of interest to both post-error slowing and the hot hand. Given the potential importance of motivation, we compared paid performance to unpaid performance in the Buckets Game. Payment was contingent on performance — participants received one point for each correct response, and higher overall scores received greater financial reward.

The post-error slowing literature suggests that participants may adopt a more cautious approach following errors, and the hot hand literature suggests that players adopt a more risky approach following success; we therefore expected, for both paid and unpaid players, post-error slowing and post-success speeding (which are equivalent results). Because the target became easier to identify over time, we expected this additional caution following errors to result in higher accuracy following errors and lower accuracy following success. Note this is the reverse of predictions based on belief in the Hot Hand. We expected financial incentives to exaggerate this post-error slowing and reversal of the hot hand. That is, we expected the performance-contingent incentives (higher financial reward for higher game scores) to enhance goal motivation and result in higher levels of cognitive control — observed as (1) overall improved performance and (2) increased post-error slowing.

As a caveat, we note that Bocskosky, Ezekowitz and Stein (2014) reported basketball players took more risk following success with little or no reduction in accuracy. Therefore, it is possible in the Buckets game that post-error slowing and post-success speeding would not be associated with any appreciable change in accuracy. This result — more difficult attempts for no loss of accuracy — would indicate an overall increase in performance following success, consistent with Bocskosky et al.'s view of the hot hand.

## 2 Method

### 2.1 Participants

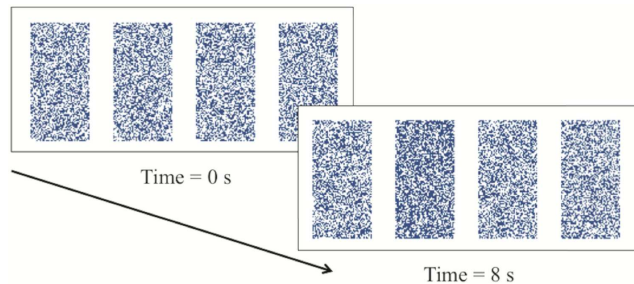
Sixty-seven undergraduates from the University of Newcastle, Australia, took part in the experiment, with 42 rewarded by course credit that was not contingent on performance. Of the 42 rewarded by course credit, 21 participated on campus in experimental testing rooms, while 21 participated online in their own time. At the beginning of the session for on-campus participants, we provided a verbal explanation of the game and encouraged them to remain motivated throughout the experiment. Despite these instructions, we found no differences between the on-campus and online sampling methods.<sup>2</sup> This is in line with findings that these two sampling methods produce equivalent results for both cognitive (Crump, McDonnell & Gureckis, 2013) and other psychological research (Gosling, Vazire, Srivastava & John, 2004). The remaining 25 players were undergraduate students, not limited to psychology, and recruited via posters placed around the campus. They participated in our testing rooms, and were paid \$10 plus 5 cents per correct target-identification, with a maximum possible payment of \$20. In addition to the standardised on-screen instructions, they received a verbal explanation of the game and the payment structure. We label the two groups in terms of reward: paid and unpaid.

### 2.2 The Buckets game

The Buckets game was coded in actionscript for Adobe Flash, an easily distributable platform that records response times with an adequate precision for our purposes (Reimers & Stewart, 2007). In the Buckets game, four 100x50 pixel rectangles ('buckets') were displayed on a computer screen, each with 50% of its pixels filled blue (blue dots). The location of blue pixels within buckets was randomly updated every 100ms, and one of the buckets was slowly filled with more blue pixels. The player was asked to identify this target (see Figure 1). The target received additional blue pixels at an average rate of 1.875 pixels per 100ms update. Players could select the target and hence terminate the trial at any time during the maximum trial duration of 80 updates (or equivalently, 8sec). A fixation-cross preceded trials and lasted 300ms. Visual (i.e., "CORRECT" or "INCORRECT") and auditory (i.e., cash register "ker ching", or incorrect buzz) feedback, lasting 500ms, was provided on the accuracy of each attempt, followed by a between-trial

<sup>2</sup>Two-tailed Bayesian independent samples t-tests, performed as per the analysis comparing paid and unpaid players below, and reported in favour of the alternate hypothesis, indicated no evidence for the hypothesis of a difference between on-campus and online sampling for post-error RT adjustments [traditional:  $BF = 1.01$ ; robust method:  $BF = 1.53$ ; matched:  $BF = 0.62$ ], or post-error accuracy adjustments [(traditional:  $BF = 0.30$ ; robust:  $BF = 0.89$ ); matched:  $BF = 0.31$ ].

Figure 1: An example of the evolution of Buckets game stimuli. Initially all buckets have the same number of blue pixels. One bucket accumulates additional blue pixels over the course of an 8s trial (unless a response was given beforehand). Note players could terminate the trial at any stage by making their selection. Additional pixels were added, and the location of pixels within each bucket was randomly updated, every 100ms. The target, 2<sup>nd</sup> from left, has been accentuated for the purposes of demonstration.



white screen for 500ms. An additional 1,650ms of between-trial white screen was applied to incorrect attempts (i.e., a time-out penalty applied to balance the reward function). If a player had not responded already, a tone briefly sounded 6,000ms after the buckets appeared to make players aware that the end of each trial was approaching.

Players undertook five time-limited blocks, each separated by enforced breaks of minimum 30s duration. The first block was a 5 min practice that did not count in the final score, and players were encouraged to use this block to explore the relative benefits of making attempts at different time points throughout a trial. The final four blocks were each 10 mins in length. The total game score was the sum of correctly identified targets over the four 10 min blocks.

On-screen instructions indicated the aim of the game was to identify as many targets as possible within the time allocated. On-screen instructions also made explicit that faster, and so more difficult responses, allowed for more attempts overall, but at a higher risk of making errors. During play, a countdown clock indicated the number of seconds remaining in the block, and a counter indicated the number of correct decisions made during the current block. Between blocks, players were provided updates on their previous block performance, and overall performance.

### 2.3 Analyses

**Post-error adjustments.** There are several methods in the literature for measuring post-error adjustments. We used three — *traditional*, *robust*, and *matched*. Each method involves calculating a difference between post-error and post-correct performance. It is useful to note that post-error slowing (PES) measured in this way can also be considered post-correct speeding. The traditional method involves subtract-

ing, for each participant, the mean RT of the post-error trials from the mean RT of the post-correct trials. Similarly for accuracy, it subtracts the conditional probability of a hit preceded by a hit from the conditional probability of a hit preceded by a miss.

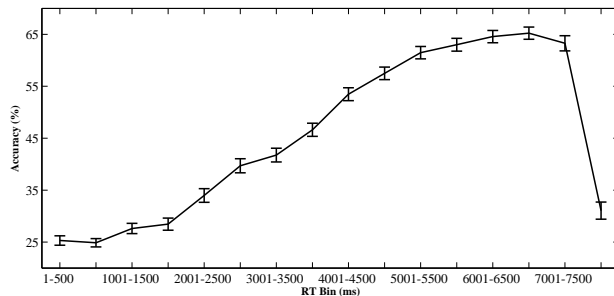
The other two measures address drawbacks of the simple global averaging used by the traditional method. One drawback is that short-term effects like post-error slowing can be confounded with long-term effects like fatigue, distraction, or boredom (Dutilh et al., 2012b). Dutilh et al. (2012b) proposed a solution that paired post-error trials with immediately preceding pre-error counterparts that are also post-correct trials. Pairwise differences are then calculated (i.e., [post-error RT] minus [pre-error, post-correct RT]), with the mean of the differences providing a robust measure of post-error RT adjustments. Dutilh and colleagues showed the robust method is able to differentiate true post-error adjustments from confounding long-term effects. Dutilh et al. (2013) employed the same type of pairs to calculate post-error accuracy adjustments. We describe these RT and accuracy based differences as “robust” measures.

A second drawback of the traditional method is that it can be confounded by systematic differences in the relative speed of correct and error responses (Hajcak & Simons, 2002). Consider a participant who slows down after all fast responses. This participant is not adjusting to errors, but is sensitive to the speed of their previous response. If errors are faster than correct responses — as is the case in the Buckets game — the traditional method of calculating post-error adjustments spuriously indicates post-error slowing. To counter such confounds, Hajcak and Simons paired each error response with a correct response closely matched on RT. We used such pairs<sup>3</sup> in the same way that pairs were used in the robust method, to calculate what we call “matched” measures based on both RT and accuracy.

**Statistical comparisons.** Null hypothesis tests cannot provide evidence in favour of the null, which is problematic because providing evidence for both null and alternate hypotheses is useful in assessing our results. Therefore, we performed all statistical comparisons using the Bayesian approach implemented in the BayesFactor package for R (Morey & Rouder, 2014), as called by JASP (Love et al., 2015) — the user friendly graphical interface for common statistical analyses. The Bayesian approach allows quantification of evidence in favor of either of the hypotheses, and each test produces a Bayes Factor (BF) that indicates

<sup>3</sup>In particular, we selected the closest matching but faster correct RT for odd errors (i.e., the first, the third, the fifth error, and so on, in terms of serial location), and the closest matching but slower correct RT for even errors. If a match within 30ms was not available this error was discarded from analysis. In the event of multiple identical RT matches, a random selection was made from those available. In the event that there were more errors than correct trials in a given data set, we began with the less common correct responses and searched for matching errors.

Figure 2: The probability of a correct decision by response time in the Buckets game. Error bars show the standard error of a proportion.



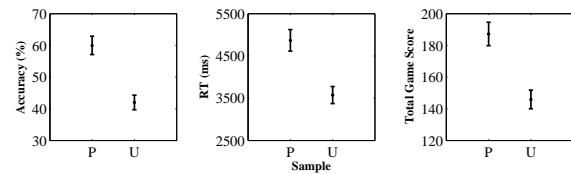
the factor by which prior beliefs should be changed by the data. We use the classification scheme proposed by Jeffreys (1961) to describe BF results. Unless otherwise specified we employed a default Cauchy prior width of  $r = 1$  for effect size, as specified by Rouder et al. (2009) and Wetzels et al. (2009).

### 3 Results

We first checked whether accuracy improved as participants waited for more dots to accumulate in the target bucket. Accuracy by time for all attempts is shown in Figure 2. Accuracy increased as expected for 0–7,000ms, but then plateaued, and dropped steeply for responses slower than 7500ms. Errors that are slower than 7500ms included both non-attempts (failure to beat the deadline) as well as incorrect attempts. Given the high error rate, we suspected the looming deadline led to late guesses. Because it was impossible to identify and separate late guesses from “proper”, non-guess attempts that resulted in errors (or hits, for that matter), we removed contributions to post-error adjustment calculations that relied on responses slower than 7500ms, which were 3% of all attempts, 25% of which were non-attempts. For example, for the robust method, if any  $e-2$ ,  $e-1$ ,  $e$ , or  $e+1$  response was slower than 7500ms, where  $e$  indicates the trial index of an error, the pre- and post-error paired difference for this quartet of trials was removed from analysis. We also removed contributions relying on responses faster than 500ms, which were 8.8% of all attempts, 44% of which come from the 2 participants who are subsequently excluded. Based on players’ self-report these very fast responses represented guesses.<sup>4</sup>

<sup>4</sup>Our findings were robust against variations in the exclusion criteria. To check, we re-ran our post-error analyses for accuracy and RT changes (as seen in Figures 4 and 5, and reported in corresponding text) for three different exclusion scenarios. Under scenario 1 no responses were excluded, and for scenario 2 only responses slower than 7500ms were excluded. For both scenarios we found an unchanged pattern of RT and accuracy post-error

Figure 3: Accuracy, mean RT, and total game score for paid (P) and unpaid (U) players. Error bars show the standard error of the mean.



We then confirmed that each participant had enough remaining responses to calculate post-error and hot-hand measures. For the traditional method we required that each participant contributed at least 20 errors and 20 correct responses. One player from the unpaid group failed to meet these criteria, having made many responses faster than 500ms. For the robust measure, we required at least 20 suitable pairs. One additional player from the unpaid group was excluded due to too many fast responses. For the matched measure, we also required 20 pairs, with no further exclusions required. This left 45 and 20 participants in the unpaid and paid groups, respectively.

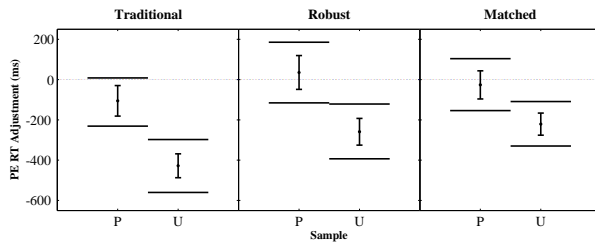
Figure 3 demonstrates the efficacy of our payment manipulation, with higher accuracy, slower responding and higher overall game scores in the paid group. We used one-sided Bayesian independent samples t-tests to quantify the evidence for the hypotheses that the paid participants would be slower, more accurate, and accumulate higher overall scores. Here we report Bayes factors in favour of the alternate hypothesis. The Bayes factors were  $BF = 271$ ,  $BF = 3780$ , and  $BF = 894$  respectively, indicating that the observed data were much more likely under the alternative hypothesis that postulates an effect of payment than under the null hypothesis that postulates the absence of the effect. This is decisive evidence in each case. We conclude that paid players were more focused on achieving the game goals than players from the unpaid group.

#### 3.1 Post-Error Analysis

**Post-error response-times.** For all methods, post-error adjustments were calculated on an individual basis. Figure 4 displays the results for post-error RT analysis and highlights two important results. Firstly, the direction of the difference between paid and unpaid participants was in line with expectations. Secondly, and surprisingly, no post-

results, and statistical reliability increased for the critical RT results. Under scenario 3 only responses faster than 500ms were excluded. Here we again found an unchanged pattern of RT and accuracy post-error results, however, the statistical reliability of RT results decreased for the traditional ( $BF = 7.55$ ) and matched ( $BF = 2.12$ ) methods, but increased for the robust method ( $BF = 20.4$ ). No participants were excluded under scenarios 1 and 2 whereas two participants were excluded under scenario 3.

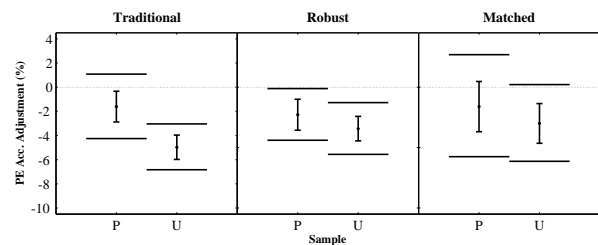
Figure 4: Post-error (PE) RT adjustment in the Buckets game for paid (P) and unpaid (U) participants, for each of the traditional, robust, and matched measurement methods. The y-axis represents post-error RT adjustment. Above zero values indicate post-error slowing, or more caution following an error. Below zero values indicate post-error speeding, or less caution following an error. The errors bars indicate the standard error of the mean. The horizontal lines indicate the 95% credible interval for the mean.



error slowing was observed for any of the groups and regardless of the method of calculation. Instead, considerable *post-error speeding* was documented for the unpaid group. This is supported by 95% credible intervals that indicate the unpaid group showed reliable post-error speeding for the traditional, robust, and matched methods. In contrast, the paid players showed no reliable post-error speeding for any method, and a near zero post-error RT adjustment for the robust and matched methods. One-sided Bayesian independent samples t-tests, reported in favour of the alternate hypotheses, confirmed that the paid sample showed less post-error speeding for the traditional, robust, and matched methods (traditional:  $BF = 43.8$ ; robust:  $BF = 10.2$ ; matched:  $BF = 3.32$ ). According to Jeffreys (1961), this is very strong, strong, and substantial evidence respectively for the alternative hypothesis that postulates payment will lead to more post-error slowing (or less post-error speeding) than under the null hypothesis that postulates the absence of the effect.

**Post-error accuracy.** Figure 5 displays results for post-error accuracy adjustments. There was a tendency for lower accuracy following errors, or equivalently, higher accuracy following success, that is, a hot hand. Overall, this tendency ranged from 2–5%, but for the paid sample this tendency was smaller and less reliable. One-sided Bayesian independent samples t-tests, reported in favour of the alternate hypothesis, confirmed that the traditional method provided anecdotal evidence for the alternate hypothesis of a larger decrease in post-error accuracy for the unpaid group ( $BF = 2.53$ ), whereas the robust ( $BF = 0.36$ ) and matched ( $BF = 0.30$ ) methods showed anecdotal and substantial evidence respectively for the null hypothesis of no difference between post-error accuracy adjustments for paid and unpaid players.

Figure 5: Post-error (PE) accuracy adjustment in the Buckets game for paid (P) and unpaid (U) players, for each of the traditional, robust, and matched measurement methods. The y-axis represents post-error accuracy adjustment. Above zero values indicate more accurate identification of the target following an error. Below zero values indicate more accurate identification of the target following success, or a hot hand. The errors bars indicate the standard error of the mean. The horizontal lines indicate the 95% credible intervals for the mean.



With the RT adjustments reported above, it seems that unpaid players become more cautious and accurate following success, or less cautious and accurate following errors.

**Short- and long term effects of errors.** Figure 4 suggested, for both paid and unpaid groups, that the traditional method indicated more post-error speeding than the other methods, for both the paid and unpaid groups. Because the traditional method captures both short-term and long-term sequential effects, whereas the other methods focus specifically on the short-term effects of errors, we used this difference to estimate the relative influence of short- and long-term effects in the Buckets game. A three (measure: traditional, robust, matched) by two (sample: paid, unpaid) Bayesian mixed model ANOVA indicated evidence for the main effects of measure and group with this two factor model maximizing the marginal probabilities relative to the null model of no effects, JASP estimating the  $BF \sim 90,000$ . While no single factor model was supported, it can be instructive to assess these models to shed light on the relative influence of the two factors. In terms of the two effects, measure had an extremely strong influence relative to the null,  $BF \sim 9,000$ , whereas group had a less pronounced effect,  $BF \sim 10$ . Bayesian paired samples t-tests, reported in favor of the alternate hypothesis and with posterior model odds calculated using model priors that were adjusted for multiple comparisons<sup>5</sup>, provided de-

<sup>5</sup>For  $k$  comparisons we set the prior probability of finding no difference in a single comparison ( $p$ ) so that the probability of finding no difference in the set of  $k$  comparisons equals the total probability of finding one or more differences. That is, we solve  $p^k = 1/2 \Rightarrow p = 2^{-1/k}$ . In the present case where  $k=3$ ,  $p=0.794$ . So if  $BF$  is the Bayes factor for Difference vs. No Difference for a particular comparison then the posterior odd (which can be conceived as a corrected Bayes Factor for the multiple comparisons) is (1

cisive evidence that the traditional method showed more post-error speeding (marginal mean =  $-267$ ms) than both the robust (marginal mean =  $-112$ ms,  $BF = 331$ ) and matched (marginal mean =  $-124$ ms,  $BF = 1657$ ) methods, and strong evidence for no difference between the robust and matched methods ( $BF = 0.03$ ). Thus, the short-term effects of errors accounted for approximately half of the post-error speeding seen in the Buckets game.

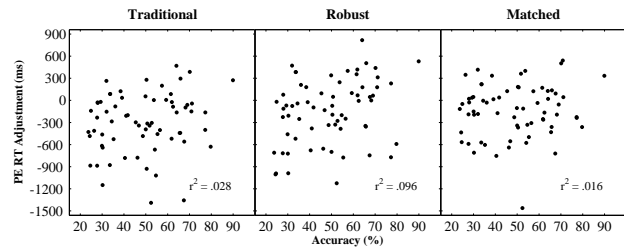
## 4 Discussion

We aimed to investigate sequential effects caused by the influence of previous-response success (or failure) on current performance. The Buckets game provided an intermediate time-scale and a carefully controlled environment so that both hot-hand and post-error statistics could be estimated from the same data. Players were either paid or unpaid, with payments structured to incentivise the Buckets game goal of maximising the number of correct target detections in a fixed time period. Past experimental investigations have typically found post-error slowing. In contrast, hot hand research has focused on professional sports settings. Although the hot hand is typically considered a fallacy (Avugos, Köppen, Czienskowski, Raab & Bar-Eli, 2013; Bar-Eli, Avugos & Raab, 2006), professional basketball players have been reported to take more difficult shots following success (Bocskocsky, Exekowitz & Stein, 2014; Rao, 2009), which would mask a hot-hand effect.

As expected, we found monetary rewards improved overall performance. We also found that financial incentives influenced post-error RT adjustments in the expected direction, toward post-error slowing. This is in line with previous findings that financial incentives improve performance in cognitive tasks (Kouneiher, Charron & Koechlin, 2009; Padmala & Pessoa, 2011; Bonner, Hastie, Sprinkle & Young, 2000; Camerer & Hogarth, 1990) and increase post-error slowing (Sturmer, Nigbur, Schact & Sommer, 2011; Ullsperger & Szymanowski, 2004). Our work provides an extension of these previous findings in that the shift we observed toward post-error slowing for paid players occurred in a novel and temporally extended task. This result was encouraging with regard to our primary theoretical investigation, it suggested that behaviour in the Buckets game — an intermediate environment between those typically used to study post-error slowing and the hot hand — showed behavioural signatures consistent with the post-error slowing literature. Thus, our data supported the broader position that increased motivation will result in an increased level of cognitive control, regardless of task (Botvinick & Braver, 2015).

p)/p X BF. For example, suppose that  $BF = 10$  (i.e., data changes our belief by a factor of 10 in favor of a difference) then the posterior odds are  $10(1/0.794)/0.794 = 2.6$ .

Figure 6: Post-error RT change by accuracy for the traditional (left), robust (middle), and matched (right) methods. R-squared indicates the proportion of variance in post-error RT changes accounted for by accuracy.



Importantly though, we found that unpaid players exhibited post-error speeding rather than slowing, and that paid participants showed neither post-error slowing nor post-error speeding. These results suggest the influence of the prior outcome may be quite different in an environment such as the Buckets game to those typically used to investigate post-error slowing or the hot hand. In regards to post-error slowing, our finding of post-error speeding for the unpaid group was especially surprising and rare.

Notebaert et al.'s (2009) *orienting account* of post-error slowing provides a potential reconciliation of this surprising result. Notebaert and colleagues proposed participants are surprised and distracted by errors when they are rare — the usual case in most post-error slowing research — and are slowed because they must reorient to the task after committing errors. Conversely, when success is rare, the orienting account predicts post-error speeding, a prediction that has been confirmed in some rapid-choice tasks when errors are more common than correct decisions (e.g., Houtman, Núñez Castellar & Notebaert 2012; Núñez Castellar, Kühn, Fias & Notebaert, 2010). Consistent with this account, in our data error rates were higher for unpaid participants (average rate of 58%) than they were for paid participants (average rate of 40%). Therefore, according to the orienting account, error rates for unpaid participants were in the region that might encourage post-error speeding, whereas error rates for paid participants were in the region that might encourage no post-error slowing.

The orienting account does not predict a difference in post-error behaviour based on the level of motivation. However, it makes another testable prediction, namely a positive relationship between the overall rate of errors and post-error RT adjustments, with post-error slowing increasing with increased accuracy. To test the orienting account we investigated the relationship between accuracy and post-error RT adjustments for all players in the two groups, paid and unpaid. Figure 6 shows that accuracy explains very little of the variance in any of the three measures of post-error RT adjustments. Bayesian tests of correlation, using a Beta prior

width of 1 and reported in favor of the alternate hypothesis, confirmed that for both the traditional ( $BF = 0.37$ ) and matched methods ( $BF = 0.25$ ), there was evidence for the null hypothesis of no relationship between accuracy rate and magnitude of post-error adjustments. For the robust method ( $BF = 3.34$ ), there was evidence in favor of a relationship. A potential reason for a lack of orienting effects is that the Buckets game had a minimum 1,300ms inter-trial-interval and a possible 8,000ms trial time. These longer time scales may have negated the impact of re-orientation. In any case, the orienting account cannot explain post-error speeding in the Buckets game.

With the orienting account excluded, a lack of post-error slowing for paid participants and the post-error speeding observed for unpaid participants suggest there are substantial differences between post-error behaviour in the Buckets game and in typical rapid-choice tasks. Future work could explore the specific task demands that are responsible for this lack of post-error slowing. To this end, it is useful to note that the Buckets game — while a novel intermediate step between rapid choice and sporting environments — is related to two other paradigms. First, it is related to the expanded judgement task developed by Irwin, Smith and Mayfield (1956), and in particular the information-controlled expanded-judgement tasks used by Brown, Steyvers and Wagenmakers (2009), and Hawkins, Brown, Steyvers and Wagenmakers, (2012). In these tasks, evidence for a target item among distractors is accumulated stochastically on screen in discrete time steps. As in the Buckets game, information toward the correct decision accumulated slowly over time, and the longer a participant waited before responding, the more likely they were to correctly identify the target. These tasks closely resemble the temporally extended nature of the Buckets game. Second, the goal driven structure of the Buckets game, in which players were asked to maximise the number of successes within a fixed time period, is related to rapid choice tasks used to investigate reward-rate optimization (e.g., Bogacz, Hu, Holmes & Cohen, 2010; Simen et al., 2009). It would be interesting, therefore, to examine post-error effects in these paradigms. In any event, a lack of post-error slowing in the Buckets game provides the empirical evidence in support of the speculations of Yeung and Summerfield (2012); there may be substantial differences in post-error behaviour for goal driven and temporally extended tasks when compared to rapid choice tasks.

Given the differences between paid and unpaid post-error performance, our data support an account of post-error speeding in the Buckets game that rests on participant motivation. We propose our unpaid participants were generally unmotivated, and rather than recruiting cognitive control, errors further decreased the level of cognitive control available. In other words, we propose that in the Buckets game environment — which we note had relatively low success rates — unpaid participants were discouraged by errors and

consequently made less cautious responses, whereas success encouraged them to try harder. This might be considered “post-error recklessness”. For the group rewarded by monetary incentive however, cognitive control was enhanced — as evidenced by better overall performance — and the discouraging impact of errors was negated, explaining why we observed post-error speeding for unpaid, but not paid, participants. It may have been that there were individual differences in the motivating effect of financial incentives, so that some paid participants were motivated to increase caution after errors, but some were discouraged by them as in the unpaid group, so that on average there was no post-error slowing. Future research might directly measure motivation in order to check whether it correlates with the level of post-error slowing. Future work may also consider whether similar mechanisms contribute to post-error speeding observed in rapid choice when error rates are very high.

With regards to the hot hand, unlike professional basketball where post-success increases in shot difficulty may mask the hot hand (Bocskocsky, Exekowitz & Stein, 2014; Rao, 2009), we found the difficulty-accuracy trade-off was most likely a major cause of us finding a hot hand effect in our unpaid players. This hot hand effect was absent for paid players. Estimated at approximately 5% by the traditional measure, our unpaid players showed a hot hand effect closer in size to that reported in hot hand beliefs (Gilovich, Vallone & Tversky, 1985) than any previous research we are aware of. This finding hints at reconciliation between hot hand beliefs and empirical data that rests on motivation and cognitive control. Specifically, when player motivation is low, a decrease in cognitive control may follow errors, and an increase in cognitive control may follow success. In this way, success breeds success. Critically, the hot hand may well remain a fallacy at the professional level when motivation is high, but fans and players may have experienced the hot hand themselves in amateur contexts where motivation is lower — hence the resilient nature of the belief. Future research might examine whether similar post-error recklessness occurs in the amateur sport context, where motivation may be lower and repeated errors may discourage players, whereas success may provide encouragement to take more care, and hence be more accurate. This would be commensurate with the findings that golfers (Cotton & Price, 2006) and tennis players (Klaassen & Magnus, 2001) with little competitive experience were more likely to demonstrate the hot hand than those with more competitive experience.

## 5 Conclusion

Our simultaneous investigation of post-error slowing and the hot hand revealed a surprising pattern of results that both supported and challenged existing theories. The take home message from our work is that caution is required when



applying our current understanding of how errors and success influence behaviour in novel contexts. We confirmed that motivation and cognitive control are central considerations when exploring the effect of previous outcomes. This conclusion is commensurate with the speculations of Yeung and Summerfield (2012); there may be substantial task-dependent differences in post-error behaviour for temporally extended and goal driven tasks. We conclude that post-error slowing should not necessarily be considered a general phenomenon in decision-making, but rather one that is pervasive in tasks that require a rapid response without much deliberation. Although replication and extension of our work to amateur sport is required, our work also has the potential to increase our understanding of hot hand beliefs. Although likely a fallacy in professional sports, the hot hand may be observed in contexts that encourage post-error recklessness.

## References

- Avugos, S., Köppen, J., Czienskowski, U., Raab, M., & Bar-Eli, M. (2013). The “hot hand” reconsidered: A meta-analytic approach. *Psychology of Sport and Exercise, 14*(1), 21–27. <http://dx.doi.org/10.1016/j.psychsport.2012.07.005>
- Bar-Eli, M., Avugos, S., & Raab, M. (2006). Twenty years of “hot hand” research: Review and critique. *Psychology of Sport and Exercise, 7*(6), 525–553. <http://dx.doi.org/10.1016/j.psychsport.2006.03.001>
- Bocskocsky, A., Ezekowitz, J., & Stein, C. (2014). *The hot hand: A new approach to an old “fallacy”*. Paper presented at the MIT Sloan Sports Analytics Conference, Boston.
- Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010). Do humans produce the speed-accuracy trade-off that maximizes reward rate? *Quarterly Journal of Experimental Psychology (Hove), 63*(5), 863–891. <http://dx.doi.org/10.1080/17470210903091643>
- Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, S. M. (2000). A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research, 12*, 19–64.
- Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annual Review of Psychology, 66*, 83–113. <http://dx.doi.org/10.1146/annurev-psych-010814-015044>
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652.
- Brown, S., Steyvers, M., & Wagenmakers, E.-J. (2009). Observing evidence accumulation during multi-alternative decisions. *Journal of Mathematical Psychology, 53*(6), 453–462. <http://dx.doi.org/10.1016/j.jmp.2009.09.002>
- Cameron, J., Banko, K. M., & Pierce, W. D. (2001). Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *The Behavior Analyst, 24*(1), 1–44.
- Camerer, C. F., & Hogarth, R. M. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*(1), 7–42. <http://dx.doi.org/10.1023/a:1007850605129>
- Cotton, C., & Price, J. (2006). The hot hand, competitive experience, and performance differences by gender. Retrieved from Social Science Research Network website: <http://dx.doi.org/10.2139/ssrn.933677>
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon’s Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE, 8*(3), e57410. <http://dx.doi.org/10.1371/journal.pone.0057410>
- Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. *Frontiers in Psychology, 2*(233), 1–9. <http://dx.doi.org/10.3389/fpsyg.2011.00233>
- Dudschig, C., & Jentsch, I. (2009). Speeding before and slowing after errors: Is it all just strategy? *Brain Research, 1296*, 56–62. <http://dx.doi.org/10.1016/j.brainres.2009.08.009>
- Dutilh, G., Forstmann, B. U., Vandekerckhove, J., & Wagenmakers, E. J. (2013). A diffusion model account of age differences in posterror slowing. *Psychology and Aging, 28*(1), 64–76. <http://dx.doi.org/10.1037/a0029875>
- Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E. J. (2012a). Testing theories of post-error slowing. *Attention, Perception, and Psychophysics, 74*(2), 454–465. <http://dx.doi.org/10.3758/s13414-011-0243-2>
- Dutilh, G., van Ravenzwaaij, D., Nieuwenhuis, S., van der Maas, H. L. J., Forstmann, B. U., & Wagenmakers, E.-J. (2012b). How to measure post-error slowing: A confound and a simple solution. *Journal of Mathematical Psychology, 56*(3), 208–216. <http://dx.doi.org/10.1016/j.jmp.2012.04.001>
- Gehring, W. J., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *The Journal of Neuroscience, 21*(23), 9430–9437.
- Gilden, D. L., & Wilson, S. G. (1995). Streaks in skilled performance. *Psychonomic Bulletin & Review, 2*(2), 260–265.
- Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*(3), 295–314. [http://dx.doi.org/10.1016/0010-0285\(85\)90010-6](http://dx.doi.org/10.1016/0010-0285(85)90010-6)
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*(2), 93–104. <http://dx.doi.org/10.1037/0003-066x.59.2.93>

- Hajcak, G., & Simons, R. F. (2002). Error-related brain activity in obsessive-compulsive undergraduates. *Psychiatry Research, 110*(1), 63–72.
- Hawkins, G., Brown, S., Steyvers, M., & Wagenmakers, E.-J. (2012). An optimal adjustment procedure to minimize experiment time in decisions with multiple alternatives. *Psychonomic Bulletin & Review, 19*(2), 339–348. <http://dx.doi.org/10.3758/s13423-012-0216-z>
- Houtman, F., Núñez Castellar, E. P., & Notebaert, W. (2012). Orienting to errors with and without immediate feedback. *Journal of Cognitive Psychology, 24*(3), 278–285.
- Irwin, F. W., Smith, W. A. S., & Mayfield, J. F. (1956). Tests of two theories of decision in an “expanded judgment” situation. *Journal of Experimental Psychology, 51*(4), 261–268. <http://dx.doi.org/10.1037/h0041911>
- Jentzsch, I., & Dudschig, C. (2009). Why do we slow down after an error? Mechanisms underlying the effects of posterror slowing. *The Quarterly Journal of Experimental Psychology, 62*(2), 209–218. <http://dx.doi.org/10.1080/17470210802240655>
- Jeffreys, H. (1961). *Theory of Probability*. Oxford: Oxford University Press.
- Klassen, F. J. G. M., & Magnus, J. R. (2001). Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association, 96*, 500–509.
- Kouneiher, F., Charron, S., & Koehlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nature Neuroscience, 12*(7), 939–945. <http://dx.doi.org/10.1038/nn.2321>
- Laming, D. (1968). *Information theory of choice-reaction times*. New York: Academic Press.
- Laming, D. (1979). Choice reaction performance following an error. *Acta Psychologica, 43*, 199–224.
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropman, D., Verhagen, A. J., ... Wagenmakers, E.-J. (2015). JASP (Version 0.7) [Computer Software].
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: An orienting account. *Cognition, 111*(2), 275–279.
- Núñez Castellar, E. P., Kühn, S., Fias, W., & Notebaert, W. (2010). Outcome expectancy and not accuracy determines posterror slowing: ERP support. *Cognitive, Affective, & Behavioral Neuroscience, 10*(2), 270–278.
- Padmala, S., & Pessoa, L. (2011). Reward reduces conflict by enhancing attentional control and biasing visual cortical processing. *Journal of Cognitive Neuroscience, 23*(11), 3419–3432. [http://dx.doi.org/10.1162/jocn\\_a\\_00011](http://dx.doi.org/10.1162/jocn_a_00011)
- Rabbitt, P. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology, 71*(2), 264–272.
- Rabbitt, P. (1966b). Error correction time without external error signals. *Nature, 212*(5060), 438–438.
- Rabbitt, P. (1969). Psychological refractory delay and response-stimulus interval duration in serial, choice-response tasks. *Acta Psychologica, 30*(0), 195–219. [http://dx.doi.org/10.1016/0001-6918\(69\)90051-1](http://dx.doi.org/10.1016/0001-6918(69)90051-1)
- Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology, 29*(4), 727–743. <http://dx.doi.org/10.1080/14640747708400645>
- Rao, J. M. (2009). *Experts’ perceptions of autocorrelation: The hot hand fallacy among professional basketball players*. <http://www.justinmrao.com/playersbeliefs.pdf>. UC San Diego.
- Read, D. (2005). Monetary incentives, what are they good for? *Journal of Economic Methodology, 12*(2), 265–276. <http://dx.doi.org/10.1080/13501780500086180>
- Reimers, S., & Stewart, N. (2007). Adobe Flash as a medium for online experimentation: A test of reaction time measurement capabilities. *Behavior Research Methods, 39*(3), 365–370. <http://dx.doi.org/10.3758/BF03193004>
- Ridderinkhof, K. R., van den Wildenberg, W. P. M., Wijnen, J., & Burle, B. (2004). Response inhibition in conflict tasks is revealed in delta plots. In M. Postner (Ed.), *Attention* (pp. 369–377). New York: Guilford Press.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review, 16*, 225–237. <http://dx.doi.org/10.3758/PBR.16.2.225>
- Morey, R. D., & Rouder, J. N. (2014). BayesFactor 0.9.6. Comprehensive R Archive Network. Retrieved from <http://cran.r-project.org/web/packages/BayesFactor/index.html>
- Schroder, H. S., & Moser, J. S. (2014). Improving the study of error monitoring with consideration of behavioral performance measures. *Frontiers in Human Neuroscience, 8*(MAR). <http://dx.doi.org/10.3389/fnhum.2014.00178>
- Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., & Cohen, J. D. (2009). Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and Performance, 35*(6), 1865–1897. <http://dx.doi.org/10.1037/a0016926>
- Smith, G. (2003). Horseshoe pitchers’ hot hands. *Psychonomic Bulletin and Review, 10*(3), 753–758.
- Sturmer, B., Nigbur, R., Schacht, A., & Sommer, W. (2011). Reward and punishment effects on error processing and conflict control. *Frontiers in Psychology, 2*, 335. <http://>

- [dx.doi.org/10.3389/fpsyg.2011.00335](http://dx.doi.org/10.3389/fpsyg.2011.00335)
- Ullsperger, M., & Szymanowski, E. (2004). ERP Correlates of error relevance. In M. Ullsperger & M. Falkstein (Eds.), *Errors, Conflicts, and the Brain. Current opinions on Performance Monitoring* (pp. 171–184). Leipzig: MPI for Human Cognitive and Brain Sciences.
- Wetzels, R., Raaijmakers, J. G. W., Jakab, E., and Wagenmakers, E.-J. (2009). How to quantify support for and against the null hypothesis: a flexible WinBUGS implementation of a default Bayesian t test. *Psychonomic Bulletin and Review*, 16, 752–760. <http://dx.doi.org/10.3758/PBR.16.4.752>
- Williams, P., Nesbitt, K., Eidels, A., Washburn, M., & Cornforth, D. (2013). Evaluating Player Strategies in the Design of a Hot Hand Game. *GSTF Journal on Computing (JoC)*, 3(2), 1–11. <http://dx.doi.org/10.7603/s40601-013-0006-0>
- Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1310–1321. <http://dx.doi.org/10.1098/rstb.2011.0416>