INFORMATION PROCESSING CONSTRAINTS AND ASSET MISPRICING*

Alasdair Brown

I analyse a series of natural quasi-experiments – centred on betting exchange data on the Wimbledon Tennis Championships – to determine whether information processing constraints are partially responsible for mispricing in asset markets. I find that the arrival of information during each match leads to substantial mispricing between two equivalent assets, and that part of this mispricing can be attributed to differences in the frequency with which the two prices are updated inplay. This suggests that information processing constraints force the periodic neglect of one of the assets, thereby causing substantial, albeit temporary, mispricing in this simple asset market.

Traders are bombarded with information on the macroeconomy, industrial sectors and on individual firms. This information comes in a variety of forms: newspaper articles, blogs, tweets, meetings, broker phone calls and colleague's emails. Even if traders are attentive and receive all of this raw information, it is inevitable that they will be unable to process it effectively, to deduce the implications for all of the assets (including potential assets) in their portfolio. In this article, I investigate whether these information processing constraints have an effect on the level of mispricing in asset markets.

I use a series of natural quasi-experiments centred on Betfair betting exchange data from the Men's Wimbledon Singles Tennis Championships of 2011 and 2012. Trading is conveniently divided between pre-match periods (when little or no information arrives) and so-called 'inplay' periods during the match (when information is arriving constantly). I hypothesise that the arrival of information means that traders' information processing constraints suddenly become binding. While before the match, bettors had time to assess and price the likelihood of, for example, a Roger Federer win and a 3-1 Roger Federer win, during the match they are unable to process the implications of new information for both these bets. Faced with this constraint, bettors may choose to update the value of the bet on Roger Federer to win rather than the bet on the specific score by which Federer will win. In other words, despite having all the necessary raw information to price both bets, information processing constraints may mean that the value of certain assets are not updated in a timely fashion, and therefore mispricing is temporarily observed.

To verify this idea, I examine the evolution of the implied win probability of each player in two markets: the win market, where the bet is explicitly priced; the set market, where the bet on the player to win is implicitly priced via a replicating portfolio. The 'replicating portfolio' compiled in the set market will yield exactly the same payoffs as

the win bet, regardless of the state of nature (outcome of the match). I find three pieces of evidence consistent with the notion that information processing constraints are a cause of asset mispricing. Firstly, the mispricing (calculated as the absolute difference between the implied win probability in the win and set markets) is substantially higher inplay than pre-match. This is where I believe that the constraint is more likely to be binding. Secondly, I use a difference-in-difference approach to assess the relative frequency of price changes in the two markets during each match. I find that the price in the set market changes much less frequently during play (even after controlling for intransient differences between the two markets and common effects of information arrival). This suggests that a proportion of the mispricing can be attributed to the price not being as regularly updated in the set market during each match. Finally, I verify that the frequent changes in price in the win market are not simply noise, by calculating the price discovery contribution of each market using a variant of the Hasbrouck (1995) methodology. Consistent with the win market becoming the preferred choice of traders when the information processing constraint becomes binding, I find that the win market contributes at least 82% of price discovery during each match.

I present further tests, the results of which are all consistent with information processing constraints as a factor in mispricing. First, I separate matches into those that are played in isolation (the semi-finals and finals of the two tournaments), and those that are, at least in part, played simultaneously with another match (those at the quarter-final stages). The idea is that any processing lag between the win and set markets – within a match – will be exacerbated by the presence of a distracting match. To put this another way, constraints on our ability to process information are more likely to bind when we are forced to process more than one signal. In line with this hypothesis, I find that mispricing is significantly higher inplay when another match is taking place at the same time. Second, I create measures to assess the uncertainty of the outcome of each match, and the importance of the most recent passage of play. I then examine how these two measures correlate with mispricing. The idea is that information of greater importance will require greater processing; if such processing is constrained during matches, we are likely to observe higher levels of mispricing at these times. Consistent with this idea, I find that mispricing is significantly higher when a match is finely poised, and immediately after important phases of play.

The implications of limited attention and limited processing power in asset markets have been modelled by, among others, Hirshleifer and Teoh (2003), Peng and Xiong (2006), Huang and Liu (2007) and Mondria (2010).[1] By identifying circumstances where traders have all the necessary information but limited processing power, my results build on the empirical work of Cohen and Lou (2012). Cohen and Lou hypothesised that conglomerate firms – with operations in a number of industrial segments – were 'complicated' in comparison to single industry firms. This is because the proportion of income generated by each industrial segment evolves over time and

---

[1] There is a large recent empirical literature linking investor (in)attention to asset prices (Barber and Odean, 2008; Corwin and Coughenour, 2008; DellaVigna and Pollet, 2009; Hirshleifer *et al.*, 2009, 2011; Louis and Sun, 2010; Da *et al.*, 2011; Chakrabarty and Moulton, 2012).

a forecast of this proportion requires analysis. The implication of this proposition is that industry-specific information would be more rapidly impounded into the price of single industry firms, and the returns of a portfolio of these firms would therefore predict the subsequent returns of conglomerate firms. This was indeed the result that they found. In other words, while investors had all the necessary raw information to price both simple and complicated firms accurately, constraints on their ability to process this information led to temporary mispricings (in their case, over the course of months).

The idea that assets which require more complicated information processing may be mispriced for longer strikes a chord with my results. A bet on a player to win by a certain score is a more precise, and perhaps more complicated, prediction than simply wagering on the same player to win the match. Snowberg and Wolfers (2010) presented betting market evidence that individuals struggle to correctly predict the likelihood of small probability events. In addition, in terms of interpreting new information received during each match, simple rules of thumb may not suffice in the set betting market. If a player wins an individual point, it is reasonable to suggest that the probability that they win the match either stays roughly the same or goes up. Winning this same point, however, may increase the likelihood of winning by a certain set score but decrease the likelihood of the other two possible set scores. Without a simple rule of thumb, deciphering the impact of each point for each of the set bets may be more difficult than conducting a similar exercise in the win market.

The trading mechanism on the betting exchange studied in this article resembles a standard limit order book on a financial exchange. Traders can post liquidity (via limit orders) or consume liquidity (via market orders). The payoffs in these assets resemble those of short-maturity zero-recovery fixed-income assets. Bettors receive a fixed amount if they are correct in their predictions, and lose their stake if they are incorrect, much as you would by investing in a bond. This is also a competitive and dynamic market, meaning that, much as in financial markets, there are significant costs to repeatedly mispricing assets. Bettors who make errors in their estimation of a player's win probability can expect to be picked off by other bettors. These three points give my analysis external validity.

There are also, it should be pointed out, advantages in using the exchange for this type of study. First, there is a clear(er) separation between zero-information periods and information periods than could be expected in any financial market. This allows me to classify the arrival of information as a treatment in this natural quasi-experiment, a treatment which pushes information processing ability towards its limit. There are spikes of information arrival in financial markets (such as during earnings announcements) but the distinction between information and zero-information periods is less stark than in sports betting. Secondly, the speed with which events unfold in these matches (over the course of hours) allows me to take a microscopic look at price changes in the presence of information. This is in contrast to the longer-horizon asset-pricing study of Cohen and Lou (2012). A third advantage is that I have a measure of mispricing that can be identified by traders in real-time and, if the mispricing breaches a certain level, can be immediately arbitraged (I discuss arbitrage on the betting exchange in more detail in Section 4). This contrasts with many *ex post* inefficiencies identified in financial markets – such as the 'post-earnings announcement drift'

(DellaVigna and Pollet, 2009; Hirshleifer *et al.,* 2009) – which are only apparent to traders once the mispricing has been and gone.

The final advantage of using betting exchange data relates specifically to my attempt to establish the role that information processing constraints play in asset mispricing. If traders update the value of one of the assets that they are trading (e.g. the win bets) – after viewing the progress of the match – this implies that they have received the necessary information to update the value of the other assets that they are trading (e.g. the set bets). Therefore, if mispricing between the two assets (during the match) is due to cognitive limitations, it should be a result of constraints on our ability to process information, rather than constraints on our ability to receive information. Whereas there is some ambiguity in earlier limited attention studies as to whether traders failed to process the new information quickly, or simply did not receive the information (DellaVigna and Pollet, 2009; Hirshleifer *et al.,* 2009; Louis and Sun, 2010), there should be no such ambiguity in this environment.

The attraction of high-frequency betting data has also encouraged the recent work of Croxson and Reade (2011) and Choi and Hui (2012), who both examine Betfair pricing during association football (soccer) matches. Croxson and Reade demonstrate that betting exchange prices incorporate the effect of goals in a timely manner. Choi and Hui, in a similar study, categorise the first goal of each match as either surprising (if scored by the underdog team) or expected (if scored by the favourite team). This approach allows for an examination of the way that prior beliefs shape reactions to information. They find that the betting exchange overreacts to highly surprising events but underreacts to less surprising events. My study is also closely related to the work of Brown (2013). Brown uses Betfair Wimbledon tennis betting data (2008–12) to compare the duration of arbitrage opportunities pre-inplay and inplay (algorithmic arbitrageurs are more likely to be active inplay). Whereas Brown focuses on the factors which determine the persistence of mispricing, my focus here is on the factors which precipitate mispricing in asset markets.

The rest of the article is organised as follows. In Section 1, I speculate on the cognitive limitations that may lead to asset mispricing. In Section 2, I describe the betting exchange and present summary statistics. In Section 3, I conduct the empirical analysis and in Section 4, I discuss alternative explanations for the pattern of results. Section 5 concludes.

1. Psychological Foundations

In this Section, I speculate on the cognitive limitations that may lead to mispricing on the betting exchange. Traders, wagering on the outcome of a tennis match, receive stimuli (in the form of the outcome of each point) throughout the match. After each point is played, I expect that they turn to analysing the implications of this point for the bets that they are trading: the win bets and the set bets. The analysis that they undertake can be thought of as two separate tasks, which in the absence of cognitive limitations would be carried out simultaneously. If they are not carried out simultaneously, we are likely to see mispricing between the two assets.

Psychologists, in a vast literature on 'dual-task interference', have noted that we are often poor at completing two tasks simultaneously, see Paschler (1994) for a

discussion. This applies even if the tasks are very simple. The original method of testing for dual-task interference was to present subjects with two stimuli. Once each stimulus was received, the subject would need to press a particular button, for example, to indicate the type of signal that had been sent. To ascertain the degree of interference between the two tasks, the experimenter would then vary the time between the two stimuli. Performance in the two tasks – measured in terms of the speed and accuracy of the subject's reaction – was found to decline as the interval between the two stimuli was reduced. (In my environment, there is only one stimulus – so the interval is effectively zero – but there are still two simultaneous tasks to complete.)

The delay in the time it takes the subject to carry out the second task, as they are arguably busy with the first task, is often referred to as the 'psychological refractory period' (PRP). One prominent explanation for the PRP is that we share processing capacity across tasks, so undertaking more than one task leads to declines in performance in terms of speed and accuracy (Kahneman, 1973). Another explanation is that we use a single processor to carry out many tasks and, therefore, if we are required to complete multiple tasks, a bottleneck forms, as tasks are placed in a queue to be completed by this single processor (Paschler, 1994). Pertinently, there is even evidence to suggest that increased similarity between the two tasks actually hinders performance (Navon and Miller, 1987). (This may be important on the betting exchange as the two tasks are very similar in character.)

Dual-task interference is the most obvious cognitive limitation in my setting but may just be one of many that has a role in asset mispricing. I am pre-supposing, for example, that traders comprehend all the necessary information on the progress of the match, and the pricing of the bets. However, psychologists have noted that there is often a substantial difference between visualisation and perception. In other words, even if an object crosses our eyeline, the information contained in this object is not necessarily comprehended. Simons and Chabris (1999), in their much-heralded gorilla experiment, found that many subjects – tasked with counting the number of basketball passes in a video – failed to spot an individual in a gorilla suit walking into the shot. This 'inattentional blindness' could have an impact on the betting exchange, where traders may fail to spot a substantial discrepancy of pricing in the win and set markets – even if the limit order book is in their view – if their focus is trained on the passage of play on court.[2]

2. The Betting Exchange

The data in my study are taken from Betfair, a betting exchange in the UK. Betfair provides a limit order book for bets on the full spectrum of sporting events, with a particular emphasis on horse-racing, football and, to a slightly lesser extent, tennis.

---

[2] Incidentally, inattentional blindness has also been identified outside of the laboratory. Chabris *et al.* (2011) staged a series of mock-fights on a college campus and asked subjects whether they had noticed anything unusual. The likelihood that subjects noticed the fights was partially determined by their attentional load at the time (i.e. whether they were set a simultaneous task which could distract them). This experiment was motivated by the case of Kenny Conley, a Boston police officer, who was convicted of perjury when he testified under oath that during the chasing of a suspect he had not witnessed the beating of another officer, despite having run straight past the attack.

The exchange operates by matching up bettors who wish to 'back' (bet on) an outcome, with those willing to 'lay' (bet against) the same outcome. Those taking the lay position are assuming the role traditionally taken by bookmakers. The exchange operates in a similar fashion to the standard financial exchanges. Specifically, bettors can either submit a market order, which matches up with an offsetting limit order in the book, or submit a limit order, which sits in the book until an offsetting market order arrives. The exchange generates revenue by charging a commission (between 2% and 5%) on the winner's profits.

I collected data on 14 matches from the quarter-final stages to the final of the Wimbledon Men's Singles Championships in 2011 and 2012 (there are seven matches in my sample from each tournament). Although this may seem a limited number of matches, high-frequency sampling means that there are just shy of 800,000 observations for some of my later regressions. Wimbledon was chosen as it is the most prestigious of the four grand-slam events in tennis, and the latter stages of the tournament were chosen as these proved the most popular matches for betting. The data were purchased from Fracsoft. For each match I randomly chose one player (to ensure independent observations) and the chosen player is disclosed later in Tables 3, 6 and 7. The data I have are time-stamped and include quoted 'back' and 'lay' prices (odds) sampled each second, for both pre-inplay periods before each match and inplay periods during each match. The data also include the last transaction price (odds) and the cumulative volume at each second of trading.

The betting on each match takes place in two markets. The first is the win market where bets are traded on the winner. The second is the set market, which allows for betting on the specific score by which each player wins. As matches in the grand-slam events are conducted on a best-of-five sets basis, the set market comprises six possible outcomes (3-0, 3-1, 3-2 to each player). I chose tennis predominantly because it has an extensive inplay period but also because it is possible to replicate the bet on a player to win with only three bets in the set market.

In all cases, I infer the implied probability of an outcome by taking the midpoint of the spread. For example, if the back odds on a player to win are O_B and the lay odds are O_L, then the implied win probability is $[1/(O_B + 1) + 1/(O_L + 1)]/2$. To calculate the corresponding implied win probability from the set market, I simply sum the implied probabilities for each of the three possible set scores by which he could win. For example, take the odds on Andy Murray to win the Final (against Roger Federer) in 2012. At the start of the match (14:10:18), the back (lay) odds on a Murray win were 2 to 1 (2.05 to 1). In other words, those who backed at this price would have received £2 (plus their stake) for each £1 they put down, in the case of a win. A bettor laying this outcome would be liable for £2.05 multiplied by the backer's stake, in the case of a Murray win, and would pocket the backer's stake otherwise. The back (lay) odds were 8.8 (9) on a 3-0 Murray win, 7.6 (7.8) on a 3-1 win and 7 (7.2) on a 3-2 win. The implied win probability is therefore 0.3306 in the win market and 0.3394 in the set market, reflecting a very small mispricing of Murray's prospects at the start of the match.

In the top two panels of Table 1, I describe the summary statistics on implied win probability for the full 14 matches. The average implied win probability in the win market (0.690867) is close to the average in the set market (0.682561). The relationship appears to be closer pre-inplay, with respective averages of 0.6928574

Table 1

Summary Statistics (Implied Win Probability)

	Observations	Mean	SD	Min.	Max.
Win market					
Pre-inplay	254,201	0.6928574	0.1988313	0.2919863	0.9569597
Inplay	138,786	0.6872213	0.2807402	0.0055	0.9852456
All	392,987	0.690867	0.2311134	0.0055	0.9852456
Set market					
Pre-inplay	254,209	0.6933364	0.1989548	0	0.963456
Inplay	142,899	0.6633922	0.3072107	0	1.507688
All	397,108	0.682561	0.2439415	0	1.507688
Mispricing					
Pre-inplay	254,201	0.0061245	0.0049588	$2.32e-06$	0.3306011
Inplay	138,786	0.0530399	0.1100309	0	0.8658171
All	392,987	0.022693	0.0692408	0	0.8658171

Notes. Summary statistics on implied win probability in the win market, implied win probability in the set market and the mispricing (absolute difference) between the two. Data are sampled both before and during each match. The data set comprises the 14 matches from the quarter-final stages onwards of the 2011 and 2012 Men's Wimbledon Tennis Championships.

and 0.6933364 (the averages above 0.5 reflect the fact that the favourite was randomly selected in the majority of the 14 matches). In Figure 1, I plot the implied win probability – of Andy Murray in the 2012 Final – as inferred from the win and set markets. Once again, the implied win probabilities track each other closely. One point to make at this stage is that there are a small number of surprising readings of implied win probability in the set market (incidentally, from matches other than the 2012 Final). For example, the maximum reading in the set market is 1.507688. Unlike in the win market, the implied win probability in the set market is not bounded between 0 and 1. While these extreme readings are rare, I do exclude them later in my study to ensure the robustness of my results.

Before I proceed, it is also worth explaining my choice of quoted prices, rather than transaction prices, to calculate mispricings. I use the midpoint of the quoted spread on the exchange for three reasons. Using transaction prices is problematic because the volume in the set market is often lower than that in the win market. As a result, I could be taking a substantially lagged valuation from the set market and wrongly inferring that there are large mispricings between the two markets. Using quoted prices, I am (as much as possible) comparing contemporaneous valuations. The second reason for not using transaction prices is the presence of 'bid-ask bounce'. Specifically, there could be a situation where the quoted prices did not change but because a back order was swiftly followed by a lay order – and these prices are separated by the spread – I could wrongly infer from transaction prices that valuations have changed substantially. By taking the midpoint of the spread, the results are not affected by the bounce caused by opposing order flow.

One other reason to use quoted prices is the famous 'favourite-longshot bias'. This is where, on average, returns from bets on favourites exceed those of bets on longshots. This bias was found as far back as Griffith (1949), and has also been observed, albeit to a lesser degree, on Betfair by Smith *et al.* (2006). The foremost explanation for the bias

(*a*)   Win Market

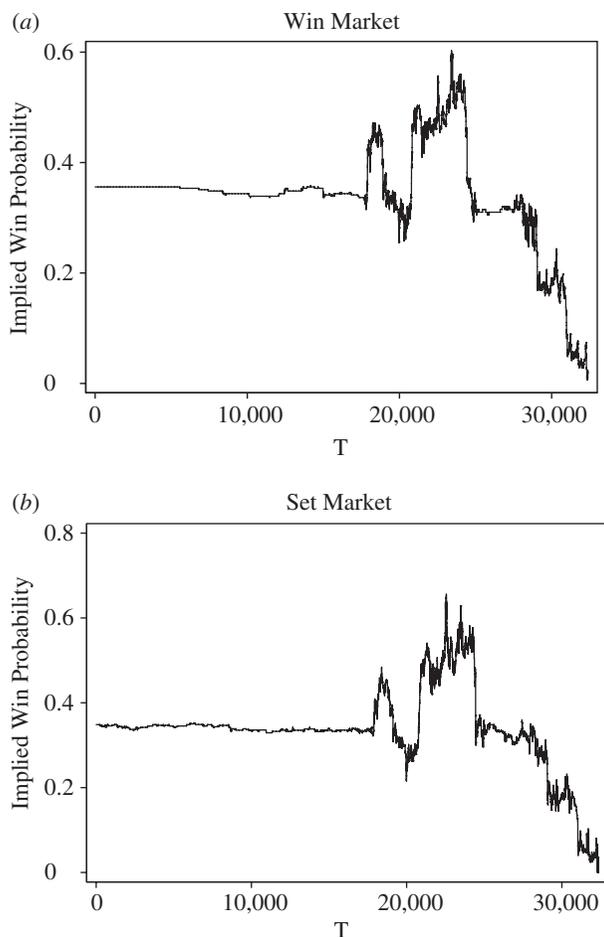(*b*)   Set Market

Fig. 1. *The Implied Win Probability of Andy Murray in the 2012 Wimbledon Men's Singles Championship Final – As Inferred from the Win Market (a) and the Set Market (b) – Plotted Against Time*
Notes. The match began at 14:10:18 (T = 17,723 on these plots).

in markets where bettors have a counterparty (including bookmaker markets and betting exchanges but excluding pari-mutuel markets) is adverse selection (Shin, 1991, 1992, 1993). Put simply, the cost to a market-maker of losing out to an insider (with advance knowledge of the outcome of the race) is greater when the winner is a longshot (as they must pay out more). The response of the market-maker is to depress the odds on the longshot further below their empirical probability than they would do for the favourite's odds. This would be a concern in my setting if I used transaction prices – which are likely to be biased towards 'backer-initiated' bets – as the set market implied win probability would regularly exceed the corresponding measure in the win market. However, using the midpoint of the quoted spread is likely to mitigate this problem, as this measure should approximate the liquidity provider's prior unbiased estimate of the probability of an outcome. Likewise, if the bias is generated by risk-loving bettors, as in Ali (1979), or over (under) estimation of small (large)

probabilities, as in Snowberg and Wolfers (2010), taking the midpoint of the spread should largely offset these effects. The fact that the average implied win probability is in fact (marginally) higher in the win market than the set market (see 'All' data in Table 1) confirms this impression.

3. Data Analysis

3.1. *Mispricing*

My first task is to establish whether the arrival of public information – in this case, the screening of a series of tennis matches – increases the level of mispricing between the win and set markets. To do this, I measure mispricing as the absolute difference between the implied win probability in the win market and the corresponding implied win probability in the set market. The bottom panel of Table 1 describes summary statistics on this measure. I find that the average mispricing between these two markets is quite small (0.02), partly because there are cases where the measure is nil. This is reassuring because, despite the differences in the way implied win probability is calculated (from 1 or 3 set(s) of prices), this demonstrates that the two markets can, at times, come to a consensus on the probability of a player's win.

Staying with the bottom panel of Table 1, we see the first evidence that mispricing is indeed higher when public information is arriving. There is a more than 10-fold increase in average mispricing during inplay periods compared to the same matches pre-inplay. Figure 2 captures this fact vividly. I plot mispricing regarding Andy Murray's implied win probability in the Final of 2012, for both pre-inplay and inplay (beginning at T = 17,723). Mispricing is visibly higher during the match.

In Table 2, I test this proposition more formally. In all of the regressions in this table, I include random effects for the 14 matches in my study, to ensure that any observed effect is widespread. In the first regression, I regress mispricing on an indicator variable
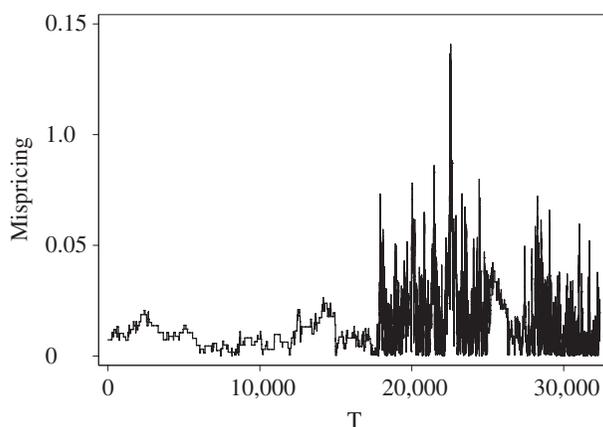

Fig. 2. *The Mispricing of Bets on Andy Murray to Win the 2012 Wimbledon Men's Singles Championship Final, Plotted Against Time*
Notes. Mispricing is measured as the absolute difference between the implied win probabilities in the win and set markets. The match began at 14:10:18 (T = 17,723 on this plot).

Table 2

Mispricing

Dependent variable: Mispricing (t)	1 All	2 Δ IWP(|) ≠ 0	3 Orders(t−1) = 0	4 IWP(Set) < 1	5 All
Intercept	0.0043192	0.0167905	0.0044452	0.0043159	0.0043245
	(0.0062537)	(0.0215517)	(0.0036837)	(0.0063836)	(0.0062911)
Inplay indicator (t)	0.0495418***	0.0429167***	0.0561157***	0.0465738***	0.0522584***
	(0.0002059)	(0.0014612)	(0.0002596)	(0.0002073)	(0.0002124)
One set outcome Indicator (t)					−0.0294461***
					(0.0005892)
ρ	0.12901166	0.51238849	0.07111057	0.1368458	0.13107971
No. of observations	392,987	41,572	266,313	385,463	392,987
R^2	0.1049	0.0127	0.1424	0.0945	0.1137

Notes. A series of regressions to compare mispricing (the absolute difference between implied win probabilities in the win & set markets) in pre-inplay and inplay periods. An indicator variable for inplay is the main explanatory variable. Regression 2 is limited to observations where the implied win probability changed (in either market) in the previous second, Regression 3 excludes observations where an order arrived (in either market) in the previous second and Regression 4 includes only observations where the implied win probability in the set market is less than 1, to exclude spurious readings of mispricing. In Regression 5, an indicator variable – equalling 1 if there is only one set outcome possible (i.e. 3-2) – is added to the specification. All 5 regressions include random effects for each match, and ρ captures the proportion of variance in the dependent variable captured by these random effects. Standard errors are in parentheses, and *, ** and *** indicate significance at the 5%, 1% and 0.1% levels respectively.

equalling 1 if the time period is during a match and 0 otherwise. I find that mispricing is higher during matches, with significance at the 0.1% level. This is a necessary but not sufficient condition for the notion that binding processing constraints are a driver of asset mispricing. The first thing to note, however, is that any measure constructed from quoted prices will undoubtedly be persistent. Therefore in regression 2, I exclude all observations where neither the implied win probability in the win market nor the implied win probability in the set market changed in the last second. This should reduce the extent to which serial correlation in mispricing is driving my result. Even after excluding these observations, I find that mispricing is significantly higher (at the 0.1% level) during matches.[3] I also wish to ensure that mispricing is not caused by the order book recovering from trade in the last period. If it were, mispricing would be a reflection of temporary illiquidity rather than differences in valuation. To deal with this issue, in regression 3, I exclude all observations where there was an order (in either the win or set market) in the preceding second. I find the aforementioned result is robust to this choice of subsample, and, in fact, the difference between the coefficients in regressions 1 and 3 suggests that trade during the match reduces rather than increases mispricing (the inplay effect is higher in regression 3). Finally, I mentioned earlier that there were a few readings of implied win probability in the set market that exceeded 1. In the fourth regression of Table 2, I exclude these readings, thereby omitting 7,520 of the 392,944 observations. Once again, I find that mispricing is higher in inplay periods, with significance at the 0.1% level.

---

[3] The result is also robust to including lagged mispricing as an additional explanatory variable, or using Newey–West standard errors, to control for persistence in the mispricing measure. The results of these two regressions are not tabulated.

The effect of information on mispricing – a more than 10-fold increase during matches judging by the coefficients in the first regression of Table 2 – is particularly striking when we consider that there are likely to be two factors acting as a restraint on mispricing. First of all, if mispricing increases beyond a certain level, arbitrage opportunities arise. An arbitrageur can construct a simple strategy of betting on a player to win in the win market, and betting against the same player in the set market (or *vice versa*). Mispricing may need to be greater than 5% for the arbitrage trade to be profitable – as Betfair commission is charged separately in the win and set market (see the discussion at the end of Section 4) – but the presence of arbitrageurs should nevertheless act as a restraint on large mispricing between the two markets. The second factor is that the progression of the match resolves uncertainty about the likely winner. At the end of the match, all uncertainty is resolved and mispricing must converge to zero. There is, after all, no more information to process.

One way of assessing the extent to which information processing constraints are responsible for this mispricing is to calculate mispricing when there is only one possible set score by which the player could win. If the player sampled either lost the match, or won by 3 sets to 2, there will be periods in which processing the implications of new information for the set bets is equivalent to processing the same information for the win bets. Essentially, in this situation the player can only win 3-2, so set bet pricing requires no additional effort to that already undertaken for the win bets. In regression 5 of Table 2, I add an indicator variable, equalling 1 if the player can only win by one possible set score at that time, and 0 otherwise. This comprises 9.13% of inplay time for the full 14 match sample. The level of mispricing is – judging by the size of the coefficients – more than 50% lower during these periods. In other words, when no additional information processing effort is required to price a set bet, the level of mispricing between win and set bets is much lower.

Up until this point, I have included random effects for each match to ensure that my results are not driven by particularly high inplay readings of mispricing in a few matches. I would like, however, to ascertain the breadth of the effect that information arrival has on asset mispricing. I do this by running the first regression of Table 2 individually for each of the 14 matches. Table 3 displays the coefficient associated with the inplay indicator for each of these matches. (I estimated these regressions with White heteroscedasticity-consistent standard errors.) The inplay effect is positive and significant in all of the 14 matches, with significance at the 0.1% level.

It is worth pausing at this stage to consider the result we would expect in this Section if information processing constraints did not bind. If information processing constraints were not binding, we would expect there to be no difference in the levels of mispricing pre-inplay and inplay. As information arrives, we would expect the implications of this information to be processed simultaneously for the two types of bet. Therefore, while the value of both the win and set bets would have changed as a result of information, there would be no effect on my measure of (relative) mispricing between the two. It is arguably only because information processing constraints bind – and therefore create a lag between the processing of information for win and set bets (I will come to this in subsection 3.3) – that we may observe higher levels of mispricing inplay.

Table 3

Mispricing (*Each Match*)

Match	Inplay indicator (t) coefficient	Standard error	N	R^2
Murray *versus* Federer 2012 final	0.0096083***	0.000139	32,384	0.1461
Tsonga *versus* Murray 2012 semi-final	0.0111361***	0.0001683	32,309	0.219
Djokovic *versus* Federer 2012 semi-final	0.0194546***	0.0002875	20,335	0.241
Federer *versus* Youzhny 2012 quarter-final	0.0399362***	0.0008172	19,681	0.2622
Djokovic *versus* Mayer 2012 quarter-final	0.0611478***	0.0008678	21,107	0.3457
Murray *versus* Ferrer 2012 quarter-final	0.032999***	0.0002968	37,880	0.3208
Tsonga *versus* Kohlschreiber 2012 quarter-final	0.3008394***	0.0027017	33,102	0.4605
Djokovic *versus* Nadal 2011 final	0.0102964***	0.0001992	19,455	0.1352
Nadal *versus* Murray 2011 semi-final	0.0188857***	0.0002306	36,794	0.2768
Djokovic *versus* Tsonga 2011 semi-final	0.0281305***	0.0004562	24,509	0.1573
Murray *versus* Lopez 2011 quarter-final	0.0255682***	0.0005666	33,394	0.1771
Nadal *versus* Fish 2011 quarter-final	0.040167***	0.000582	34,287	0.2547
Djokovic *versus* Tomic 2011 quarter-final	0.0503126***	0.0004834	22,938	0.3924
Federer *versus* Tsonga 2011 quarter-final	0.026002***	0.0003781	24,812	0.1846

Notes. A repeat of regression 1 of Table 2, this time run individually for each of the 14 matches in the data set. The only coefficient displayed is that associated with the indicator variable for inplay periods. In each match, bets on only one player (the first listed) were sampled. All regressions are estimated with White heteroscedasticity-consistent standard errors (in parentheses), and *, ** and *** indicate significance at the 5%, 1% and 0.1% levels respectively.

Nevertheless, it is important to distinguish between mispricing that arises as a result of the arrival of information, and mispricing that arises due to constraints on our ability to process this information. To this end, I am inspired by the work of Hirshleifer *et al.* (2009). These authors found that information from earnings announcements was less likely to be incorporated into asset prices when there were a number of announcements competing for the same investors' limited attention on the same day. With this in mind, it is possible that the level of mispricing that we observe on the betting exchange is affected by whether other matches are taking place on the same day. If there are other matches taking place, they may act as distractions which exacerbate the processing lag between the two types of bet. In 2011 and 2012, the quarter-finals were staged on a Wednesday, with the semi-finals staged on a Friday and the final on a Sunday. With four quarter-final matches taking place on the same day, two are staged on Centre Court (the prime venue), with the other two on Court One. This means that for much of the quarter-final stage, another prominent match acts as a distraction. The semi-finals, on the other hand, are staged sequentially on the same court and are therefore not subject to distracting events (at least not in the men's tournament), and neither is the final.

Following this intuition, in regression 2 of Table 4, I add an indicator variable for whether the match was a semi-final or final (i.e. took place in isolation), and I also interact this indicator with the inplay indicator variable (Regression 1 from Table 2 is also recreated in Table 4 for comparison purposes and random effects for each match are used, as before). I expect that if the match is played in isolation, the level of mispricing should be lower (particularly inplay), as traders can concentrate on

Table 4

Mispricing (*Further Analysis*)

Dependent variable: Mispricing (t)	1 All	2 All	3 All	4 All
Intercept	0.0043192	0.0041811	0.0472189***	0.0469867***
	(0.0062537)	(0.0084557)	(0.0083643)	(0.0063021)
Inplay indicator (t)	0.0495418***	0.0758816***	0.0223086***	0.0218303***
	(0.0002059)	(0.0002678)	(0.0005582)	(0.0005571)
SF onwards indicator		0.0019596	0.0459945***	0.0457631***
		(0.0129165)	(0.0127642)	(0.0096099)
SF onwards indicator × inplay indicator		−0.0598819***	−0.1048588***	−0.1046094***
		(0.0004037)	(0.0005888)	(0.0005875)
Uncertainty			−0.2497721***	−0.2484753***
			(0.0029973)	(0.0029906)
Uncertainty × inplay indicator			0.2959461***	0.2896722***
			(0.0026048)	(0.0026034)
abs[Δ in IWP(win)]				0.7407831
				(0.6000655)
abs[Δ in IWP(win)] × inplay indicator (t)				1.491956*
				(0.6023922)
ρ	0.12901166	0.14046007	0.14175178	0.08582684
No. of observations	392,987	392,987	392,987	392,945
R^2	0.1049	0.164	0.1379	0.1421

Notes. Further analysis related to mispricing, defined as the absolute difference between the implied win probabilities in the win and set markets. Mispricing is regressed on an inplay indicator (as in Table 2), an indicator equalling 1 if the match was a semi-final or final in either of the years (i.e. took place in isolation), a measure of the uncertainty of the match (defined as 0.5-abs[IWP(win)-0.5]), and a measure of the magnitude of information that has just arrived (defined as abs[Δ in IWP(win)]). Interactions of these variables with the inplay indicator are added to establish the specific role of the these effects, on mispricing, during the matches. All 4 regressions include random effects for each match, and ρ captures the proportion of variance in the dependent variable captured by these random effects. Standard errors are in parentheses, and *, ** and *** indicate significance at the 5%, 1% and 0.1% levels respectively.

processing the progress of one match for the various types of bets available. Indeed, this is the result I find, with mispricing significantly lower inplay in the latter stages of the tournament compared to the quarter-finals.

One caveat at this stage is that there may be confounding differences between the quarter-final stages and later rounds. Even though I have interacted the 'semi-final onwards indicator' with the inplay indicator (so such a confound would have to materialise only inplay), it must be acknowledged that Wimbledon is a seeded tournament, so we can expect that the latter stages will be contested by more evenly matched players. To capture this potential confound, I constructed a measure of 'uncertainty' in the outcome of the match, defined as 0.5 minus the absolute difference between 0.5 and the implied win probability (in the win market). (Summary statistics in Table 1, and the forthcoming analysis in subsection 3.3, attest to the greater reliability/efficiency of the win market prices.) This uncertainty measure varies between 0 (when one player is nigh-on certain to win) and 0.5 (when the match is perfectly balanced). When I add this measure (and its interaction with the inplay indicator) to my analysis in regression 3 of Table 4, I find that I may actually have been underestimating the effect that distracting events (i.e. other simultaneous matches) have on the level of mispricing. The level of mispricing is even further reduced for

matches played in isolation – compared to regression 2 – when I control for the uncertainty of the match. From regression 3, I can observe that the more uncertain the match is (specifically while it is inplay), the greater the level of mispricing. This is presumably because information (such as the outcome of each point/game), takes on greater importance when the players are locked in a tight battle and, as the importance of each point increases, so do the issues that arise in processing this information for both win and set bets.

Along the same lines, we would expect that the level of mispricing observed inplay should be affected by the importance of the most recent passage of play. Not all passages of play are equally important. There are inevitable lulls in play, such as when players sit down between games, during rain-breaks, and during sets where one player has taken a large (almost) unassailable lead. We would expect mispricing to be lower here – as there is little or no information to process – than it is during a tight tie-break, for example. In regression 4, I add a measure of the importance of the most recent piece of information, defined as the absolute change in the implied win probability in the win market, in the last second. (As mentioned before, my overall analysis suggests that the win market is the more efficient of the two, so provides a more reliable measure of the importance of the information that has just arrived.) When this measure is added, along with its interaction with the inplay indicator, I find that the level of mispricing is indeed higher when important information has just arrived (with significance at the 5% level) and that this effect is specifically found inplay (i.e. when the vast majority of information arrives).[4]

To conclude at this stage, information does appear to induce mispricing on the betting exchange. In addition, mispricing is found to be higher when there are distracting events, and when the match is finely poised. It must be re-stated, however, that higher levels of mispricing during matches are only a part of the jigsaw. In the next Section, I examine each price change – in both markets – in an attempt to establish further the role that information processing constraints play in the formation of mispricing during matches.

3.2. *Information Processing Constraints*

To examine whether information processing constraints can explain mispricing, I set up a difference-in-difference model. The aim is to assess the frequency of any changes in valuations during each match. The idea behind the regressions that follow is that I must, first, control for intransient differences between the win and set market (in terms of volume, prominence, price mechanisms etc.), and, second, I must also control for common effects that the arrival of information has on both markets. Once I have controlled for these two factors, I can then isolate the different impact that information has on the frequency of reaction in each market.

---

[4] Using absolute changes in the win price (in the last second) is, of course, not the cleanest way to measure the importance of the most recent information. The ideal technique would be to exogenously time-stamp breaks in play (such as during rain, or between games). However, video footage (where available) would only allow for time-stamping to the nearest minute (when observing the clock on the scoreboard), and internet descriptions of the game (such as on the BBC website) are also only time-stamped to the nearest minute. This precision compares unfavourably to the precision of the betting data which are observed every second.

In the top panel of Table 5, I regress an indicator variable equalling 1 if the implied win probability changed in the last second, and 0 otherwise, on three explanatory variables. The first explanatory variable is an indicator for whether the market is the win market (to control for intransient differences between the two markets), the second variable is an indicator for whether the match is inplay (to control for the common effect of information on the frequency of price revision) and the third variable is an interaction between the two aforementioned indicators. The interaction term is crucial as this captures any differences in the frequency of the two markets' responses during information arrival. A logit specification is used and random effects for each match are included. There are three results, all significant at the 0.1% level. First, in the baseline pre-match (no information) period, it appears that the set market is more susceptible to changes in valuations as the coefficient associated with the win market indicator is negative. This is perhaps expected, as trade in the set market allows for finer distinctions between different implied win probabilities as this measure is constructed from three prices rather than just one. Second, both markets respond more frequently when information arrives as the coefficient associated with the inplay indicator is positive. This is certainly expected, as traders are more likely to revise their valuations when new information arrives. Third, judging by the coefficient associated with the interaction term, it is the win market that is more likely to respond during matches. This evidence is consistent with the notion that the arrival of information means that constraints on traders' capacity to process information suddenly become binding. Rather than interpret the implications of the last point for the eight bets in the two markets (two in the win market and six in the set market),

Table 5

Information Processing Constraints

Dependent variable: Change in implied win probability (Indicator)	All	If orders$(t-1) = 0$
Intercept	−4.520144***	−5.579031***
	(0.0675464)	(0.0800766)
Win market indicator	−1.572785***	−2.175892***
	(0.0461928)	(0.1063822)
Inplay indicator $(t-1)$	2.755914***	3.364475***
	(0.0207224)	(0.0347103)
Win market indicator × inplay indicator $(t-1)$	1.729583***	2.294867***
	(0.0473131)	(0.1077369)
ρ	0.0175126	0.0220511
No. of observations where dependent variable = 1	45,145	15,914
No. of observations	794,214	650,456

Notes. Two regressions to assess whether the arrival of information inplay causes information processing constraints to bind and therefore means that the price in one market is updated less frequently. The dependent variable is an indicator variable equalling 1 if the implied probability changed in the last second. The explanatory variables are an indicator for the win market, an indicator for inplay periods and an interaction between these two indicators. The interaction term is crucial, as it captures the differences in the frequency with which the two markets respond to information (the treatment). In the second regression, I exclude all observations which immediately follow an order in the market concerned. Both regressions incorporate random effects for each match, and ρ captures the proportion of variance in the dependent variable that can be attributed to these random effects. Standard errors are in parentheses, and *, ** and *** indicate significance at the 5%, 1% and 0.1% levels respectively.

they are only able to process information related to the win bets. When there is a pause inplay – for example, when players sit down at the change of ends – perhaps they are then able to update their valuations of all the set betting outcomes. This would explain why updating in the set market is much less frequent inplay.

Crucially, this difference-in-difference analysis allows me to separate out the effect of information arrival from the effect of binding information processing constraints. The indicator for the inplay period controls for the frequency of prices changes as information arrives. The interaction term then captures the difference in the frequency with which the two markets respond to such information. If constraints were not binding, there should be no significance attached to this interaction term. The fact that there is gives information processing constraints a key role in the mispricing I documented in the previous subsection.

Returning to analysis, one concern at this point is that changes in valuations may be a reflection of asymmetries in order flow. I mentioned earlier that volume in the win market was typically higher than volume in the set market. In order to ensure that changes in price are reflections of changes in traders' valuations, rather than a result of the transient impact of orders, I also repeated the first regression in Table 5 but this time excluded all observations where an order had taken place in the market concerned in the previous second. As a result, I can focus solely on changes in valuation that are not induced (directly, at least) by orders. These results are presented in regression 2 of Table 5. The effect remains – the set market responds less frequently than the win market – and indeed is stronger for this choice of subsample.

Although I have used random effects to incorporate factors idiosyncratic to each match, in Table 6 I repeat my two regressions individually for the 14 matches in the sample. In each case, I display the coefficient associated with the interaction between the win market and inplay indicators, in order to capture the relative frequency of price changes during each match. The results that correspond to the first (second) regression in Table 5 are displayed in the top (bottom) panel of Table 6. (All regressions in Table 6 are estimated using White heteroscedasticity-consistent standard errors). My earlier results are replicated in all of the 14 matches for the first regression and 12 of the 14 matches in the second regression. In each case where statistical significance (at least at the 5% level) is found, the win market responds more often to information inplay.

3.3. *Price Discovery*

At this stage, I have provided two sets of results that are consistent with the notion that information processing constraints are partially responsible for mispricing in this market. The first piece of evidence was that mispricing was higher inplay (when constraints were more likely to be binding) and the second piece of evidence was that this was partially driven by differences in the frequencies with which prices were updated in the two markets examined. Essentially, I could argue that the set market valuation differed from that of the win market inplay because information processing capacity was focused on the latter market.

Table 6

Information Processing Constraints (Each Match)

Match	Win market indicator×inplay indicator ($t-1$) coefficient	Standard error	N	Pseudo-R^2
All				
Murray *versus* Federer 2012 final	1.497584***	0.1273677	64,766	0.2091
Tsonga *versus* Murray 2012 semi-final	2.94272***	0.1625545	64,636	0.2366
Djokovic *versus* Federer 2012 semi-final	1.689284***	0.182125	40,670	0.2271
Federer *versus* Youzhny 2012 quarter-final	3.830158***	0.2578386	43,006	0.2872
Djokovic *versus* Mayer 2012 quarter-final	4.192396***	0.5858891	44,428	0.2338
Murray *versus* Ferrer 2012 quarter-final	0.9169568***	0.1149961	75,766	0.1953
Tsonga *versus* Kohlschreiber 2012 quarter-final	2.026934***	0.1691004	66,744	0.2909
Djokovic *versus* Nadal 2011 final	0.4645506*	0.1888983	38,912	0.2159
Nadal *versus* Murray 2011 semi-final	0.8167729***	0.1403868	73,886	0.2403
Djokovic *versus* Tsonga 2011 semi-final	1.697413***	0.2571926	49,238	0.1963
Murray *versus* Lopez 2011 quarter-final	1.723645***	0.2436261	66,878	0.2179
Nadal *versus* Fish 2011 quarter-final	3.239931***	0.585237	69,754	0.2405
Djokovic *versus* Tomic 2011 quarter-final	1.596654***	0.2563132	45,906	0.184
Federer *versus* Tsonga 2011 quarter-final	3.162542***	0.415702	49,624	0.17
If orders($t-1$) = 0				
Murray *versus* Federer 2012 final	2.380325***	0.430861	47,570	0.2205
Tsonga *versus* Murray 2012 semi-final	2.947642***	0.3534875	53,803	0.2342
Djokovic *versus* Federer 2012 semi-final	3.114907***	0.7344183	29,971	0.247
Federer *versus* Youzhny 2012 quarter-final	5.29763***	0.591171	39,314	0.382
Djokovic *versus* Mayer 2012 quarter-final	3.915712***	0.5929925	40,813	0.2713
Murray *versus* Ferrer 2012 quarter-final	0.9882271***	0.2492407	59,502	0.2297
Tsonga *versus* Kohlschreiber 2012 quarter-final	2.266875***	0.290828	60,004	0.2741
Djokovic *versus* Nadal 2011 final	0.039718	0.4784036	25,184	0.2658
Nadal *versus* Murray 2011 semi-final	0.9270525**	0.3057099	57,507	0.2944
Djokovic *versus* Tsonga 2011 semi-final	1.874459*	0.7271638	36,091	0.2553
Murray *versus* Lopez 2011 quarter-final	2.34514**	0.7233241	60,173	0.2612
Nadal *versus* Fish 2011 quarter-final	3.570462***	1.012624	62,924	0.2601
Djokovic *versus* Tomic 2011 quarter-final	0.5227171	0.396602	37,472	0.2206
Federer *versus* Tsonga 2011 quarter-final	2.085963***	0.7219557	40,128	0.2131

Notes. A repeat of the two regressions in Table 5, this time estimated for each match in the sample. The only coefficient displayed is that associated with the interaction term between the win market indicator and the inplay indicator. This interaction term captures the relative response of the two markets to the treatment (information). In each match, bets on only one player (the first listed) were sampled. All regressions are estimated with White heteroscedasticity-consistent standard errors (in parentheses), and *, ** and *** indicates significance at the 5%, 1% and 0.1% levels respectively.

One gap in this argument, however, is that I cannot be sure at this stage that the frequent price changes in the win market are not simply noise. If limited processing capacity is truly being focused on the win market bets, then we should see that win market price changes lead set market price changes. In other words, only when there is a lull in play (or indeed only when arbitrageurs arrive on the scene) does the set market catch up with the information processing that has already taken place for the win bets.

To verify this idea, I use a variant of the price discovery models of Garbade and Silber (1983) and Hasbrouck (1995). As my setting has only two prices, rather than n as in Hasbrouck (1995), my model more closely resembles that of Garbade and Silber (1983). I define W_t as the implied probability of a player winning in the win market, and S_t as the implied probability of the same player winning in the set market. Both implied win probabilities are defined in Section 2. I then estimate the following two regressions:

$$W_t - W_{t-1} = \beta_0 + \beta_1(S_{t-1} - W_{t-1}) + \epsilon_t \tag{1}$$

$$S_t - S_{t-1} = \alpha_0 + \alpha_1(W_{t-1} - S_{t-1}) + u_t \tag{2}$$

where ϵ_t and u_t are error terms. The idea behind these regressions is straightforward. The greater the coefficient β_1, the greater the contribution of the set market to price discovery. A high β_1 would suggest that a mispricing between the two markets at time $t-1$ is corrected by a subsequent price change in the win market $W_t - W_{t-1}$. This would imply that the set market leads the win market in terms of price discovery, and is the location of the initial information processing. On the flipside, however, the greater the coefficient α_1, the greater the price discovery contribution of the win market. A mispricing at time $t-1$ is corrected by a subsequent price change in the set market $S_t - S_{t-1}$. I expect that $\alpha_1 > \beta_1$, as the win market is the more frequent updater of price (see subsection 3.2).[5]

Using these estimated coefficients, I calculated the win market contribution to price discovery, defined as $[\alpha_1/(\beta_1 + \alpha_1)]$. This measure, along with the coefficients for each of the 14 matches, is displayed in Table 7. (All regressions in Table 7 are estimated using White heteroscedasticity-consistent standard errors.) Sampling for this data is carried out at 1 minute intervals to allow sufficient time for information to arrive (this leaves time for approximately one point to be played). In line with my hypothesis – that information processing constraints bind inplay and that limited processing power is focused on the win market – I find that the win market is the major (and sometimes sole) contributor to price discovery. The lowest win market contribution (82%) is found in the first semi-final of 2012. In some cases, the contribution of the set market is actually negative, implying that on the occasions that the set market leads the win market, it is more often wrong than right (i.e. win market prices subsequently go in the other direction). This provides quite clear evidence that information processing capacity is being focused on the win market. I checked the robustness of my results to varying the sampling interval (10 seconds, 30 seconds, 2 minutes) and also used

---

[5] For more details on price discovery models, as applied to betting exchange data, see Chapter 2 of Brown (2012).

Table 7

Price Discovery Contribution

Match	β_1	α_1	Win market contribution	N
Murray *versus* Federer 2012 final	0.0233286	0.3077421**	0.93	244
Tsonga *versus* Murray 2012 semi-final	0.1312763	0.6024121***	0.82	139
Djokovic *versus* Federer 2012 semi-final	−0.0156917	1.020275***	1.02	167
Federer *versus* Youzhny 2012 quarter-final	0.0171214.	0.7204173**	0.98	83
Djokovic *versus* Mayer 2012 quarter-final	0.0075639	0.486052***	0.98	107
Murray *versus* Ferrer 2012 quarter-final	−0.0247664	0.510142***	1.05	257
Tsonga *versus* Kohlschreiber 2012 quarter-final	0.0024792	0.1484792**	0.98	169
Djokovic *versus* Nadal 2011 final	−0.2852521*	1.043141***	1.38	149
Nadal *versus* Murray 2011 semi-final	0.0502479	1.090315***	0.96	178
Djokovic *versus* Tsonga 2011 semi-final	−0.0778784*	0.8453572**	1.10	185
Murray *versus* Lopez 2011 quarter-final	0.0271175**	0.5637822**	0.95	122
Nadal *versus* Fish 2011 quarter-final	0.0028727	0.743369**	1.00	164
Djokovic *versus* Tomic 2011 quarter-final	0.0058966	0.5551947***	0.99	161
Federer *versus* Tsonga 2011 quarter-final	−0.0900856*	0.7017461***	1.15	189

Notes. A Table describing the win market contribution to price discovery – defined as $\alpha_1/(\beta_1 + \alpha_1)$ – for each of the 14 matches in the data set. β_1 and α_1 are estimates from (1) and (2) respectively. Data are sampled at 1 minute intervals and, in each match, bets on only one player (the first listed) were sampled. Price discovery contributions of greater than 1 indicate that on the occasions that the set market leads the win market, it is more often wrong than right (i.e. win market prices subsequently go in the other direction). All equations were estimated with White heteroscedasticity-consistent standard errors, and *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively.

Newey–West standard errors to account for serial correlation. The results are qualitatively the same so I do not present them here.

4. Discussion

In this Section, I will discuss two alternative explanations for the pattern of results described in this article. The first alternative explanation is the 'gradual information flow' hypothesis (developed by Hong and Stein, 1999 with empirical evidence in Hong *et al.,* 2000). The idea behind this model is that information – particularly private information – diffuses gradually across a population. This creates price momentum as the implications of a piece of new information are reflected in asset prices only after a significant lag. On the face of it, this does not seem an appropriate model for sports betting information. Bettors all view the same match – on television or at the stadium – and therefore can expect to be privy to the same information at the same time. This may not, however, be true. Television pictures of sporting events, including tennis, are broadcast with a lag. This means that those watching the tennis at the venue itself will receive each public signal a few seconds before those watching the same match on television. There are rumours that a number of Betfair customers, so-called 'courtsiders', have been exploiting this opportunity by placing bets with advance knowledge of the outcome of the last point (*Guardian* 29 June 2011). If 'courtsiders' are concentrated in the win market, this could explain why the win market leads the set market in price discovery.

The problem with this explanation for the results is that it is consistent with only two of the three results presented in this article. It is consistent with greater

mispricing of bets inplay (in subsection 3.1), as information on the progress of the match diffuses gradually across the two markets. It is also consistent with the win market leading the set market in subsection 3.3 (if 'courtsiders' do indeed concentrate in the win market). It is not, however, consistent with the quasi-experimental results in subsection 3.2. I find that the set market price is not simply updated later, but it is also updated less frequently. This is consistent, to my mind, only with the idea that information processing constraints force the periodic neglect of the set market.

A second possible explanation is that different interpretations of public information are driving my results. Harris and Raviv (1993) and Kandel and Pearson (1995) present models where traders with different likelihood functions differ in their interpretation of public information. This explanation may be consistent with the mispricing effect in subsection 3.1, as information arrival could induce disagreement between the two markets (if they have slightly different trading populations). It is not consistent, however, with the other two results. There is no reason why different interpretations should lead to differences in the frequencies with which traders respond to information (unless one set of traders is utterly unresponsive to certain information). In addition, different interpretations cannot create the price discovery relationship documented in subsection 3.3 unless traders learn from the beliefs of others. This learning is not permitted in the models of Harris and Raviv (1993) and Kandel and Pearson (1995) as otherwise disagreement would immediately disappear.

The results in this article do leave some open questions, however. Why, when algorithmic trading is commonplace in asset markets, is human information processing still required to eliminate mispricing? Why, after a human has interpreted the implications of the last point for the win bets, does an automated trader not instantaneously bring the set market in-line by a process of arbitrage?

First, in answer to these questions, there is some evidence that mispricings are in fact eliminated in a reasonably short period of time. Returning to Figure 2, we can observe from the number of short, sharp spikes that mispricings come and go quite rapidly. (Of course, not so rapidly as to be unobservable to the econometrician.) It seems, however, that just as one mispricing disappears, it is replaced by another, as new information arrives frequently during the match.

A second point relates to the limits of arbitrage (Gromb and Vayanos, 2010) and the limits of computerised trading. Betfair allows for algorithmic trading through their Application Programming Interface. This allows automated traders to bypass the conventional website and access the limit order book directly, thus hastening the trading process. Nevertheless, it is not possible for algorithmic traders to eliminate all of the mispricings that we observe, at least not in a risk-free manner. As mentioned earlier, Betfair charges a commission (between 2% and 5%, dependent on the volume of each trader's prior betting activity). This commission is payable on the net winnings in both the win and the set market.

To see why the Betfair commission may lead to persistent (and therefore easily observable) mispricing, I display two panels in Figure 3. The top panel contains a shortened version of Figure 2, displaying the mispricing of bets on Andy Murray in the Final of 2012, up until the point at which he lost a set. The bottom panel of Figure 3
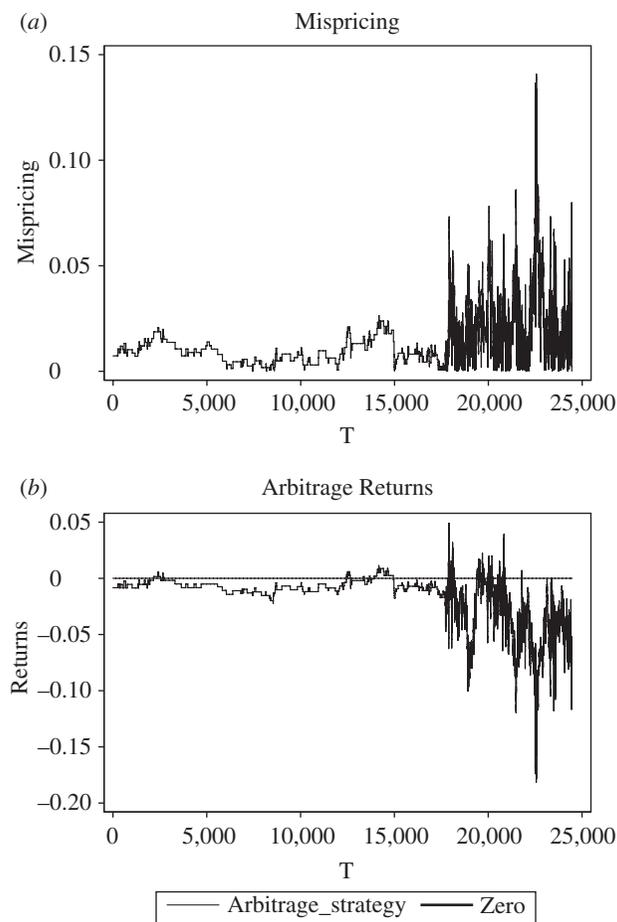
(*a*)   Mispricing

(*b*)   Arbitrage Returns

Fig. 3. *(a) The Mispricing of Bets on Andy Murray to Win the 2012 Wimbledon Men's Singles Championship Final, Plotted Against Time. Mispricing is Defined as the Absolute Difference Between the Implied Win Probabilities in the Win and Set Markets. (b) The Returns to a Risk-free Arbitrage Strategy – as Described in Brown (2013) – Which Takes a Short Position in the Win Market and a Long Position in the Set Market*
Notes. A commission level of 2% (the lowest possible) is assumed for this strategy. Data are displayed until Murray lost a set.

displays the available returns to an arbitrage strategy over the same period (assuming a commission level of 2%). The arbitrage returns data are taken from Brown (2013). The strategy involves a short position in the win market and a long position on all of the three possible set outcomes (incidentally, the opposite strategy is even less profitable over this period). This strategy could be implemented algorithmically and is coded – using the method of Brown (2013) – so as to yield an identical return irrespective of the outcome of the match.

The natural conclusion from this graph is that arbitrage opportunities are observed less often than simple mispricings (even though I assume a commission

rate as low as 2%).[6] Until mispricing breaches a certain level, traders cannot arbitrage this discrepancy remotely. This may explain why liquidity providers leave set quotes that will soon be out of date, knowing that they are unlikely to be picked off. It may also explain why we must often wait for lulls in play – and, in some cases, for the end of the match and the resolution of all uncertainty – for mispricing to totally disappear.

5. Conclusion

Limited attention has become a feature of models of investor learning behaviour (Peng and Xiong, 2006) and portfolio allocation (Huang and Liu, 2007; Mondria, 2010). Some of the more prominent empirical puzzles and anomalies in financial markets – such as stock price momentum and the underdiversification of investor portfolios – can be generated in these models. Moreover, these propositions rely on an extremely uncontroversial assumption: traders do not have unlimited capacity to receive and process information.

In this article, I examine the evidence that information processing constraints lead to asset mispricing. In the process of my examination, I exploit the unique conditions present on a UK sports betting exchange. Assets – contingent on the outcome of a tennis match – are traded in two markets (the win and the set market) both pre-inplay (when no information arrives) and inplay (when information is turned on like a tap). I argue that the arrival of information means that traders' information processing constraints suddenly become binding.

I then present three pieces of evidence consistent with the notion that information processing constraints are a cause of mispricing. The level of mispricing between the two assets is 10 times greater during the arrival of information compared to zero-information periods. This is the time when information processing constraints are more likely to be binding; part of this mispricing can be attributed to differences in the frequency with which traders update the values of the two assets during each match; price discovery is led by the market with the most frequent updating of prices (during the treatment). This suggests that traders' limited information processing capacity was predominantly put to work in the pricing of just one asset. Put together, these three results suggest a role for traders' information processing constraints in the formation of mispricing in asset markets.

The environment that I study is relatively simple. All of the assets are short lived and have binary payoffs. Although information arrives at high-frequency, it is arguably quite simple to factor the binary outcome of each point into an updated asset price. Furthermore, mispricing is tethered by the presence of a replicating portfolio – which allows for risk-free arbitrage between markets – and the imminent end of each match when all uncertainty is resolved.

Contrast this environment with that found in the stock market, for example, where the distribution of payoffs is (almost) continuous, information arrives in complex

---

[6] Incidentally, if you consider the existence of arbitrage opportunities to be a purer measure of mispricing than the measure used in this article, arbitrage opportunities are also statistically more likely to arise inplay – than pre-inplay – across the full 14 matches.

quantitative and qualitative forms, the pricing of assets requires the discounting of cash flows that arrive far into the future, and where the level of mispricing is often untethered (Stein, 2009). If I find evidence to suggest that information processing constraints cause mispricing in this simple betting market, the potential for such an effect in more complicated financial markets, is vast.

University of East Anglia

Additional Supporting Information may be found in the online version of this article:

Data sets.

References

Ali, M.M. (1979). 'Some evidence on the efficiency of a speculative market', *Econometrica*, vol. 47(2), pp. 387–92.

Barber, B.M. and Odean, T. (2008). 'All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors', *Review of Financial Studies*, vol. 21(2), pp. 785–818.

Brown, A. (2012). 'Essays on the economics of betting markets', SOAS, University of London, PhD Thesis.

Brown, A. (2013). 'Information acquisition in ostensibly efficient markets', Applied and Financial Economics Working Paper Series No. 43, University of East Anglia.

Chabris, C.F., Weinberger, A., Fontaine, M. and Simons, D.J. (2011). 'You do not talk about Fight Club if you do not notice Fight Club: inattentional blindness for a simulated real-world assault', *i-Perception*, vol. 2 (2), pp. 150–3.

Chakrabarty, B. and Moulton, P.C. (2012). 'Earnings announcements and attention constraints: the role of market design', *Journal of Accounting and Economics*, vol. 53(3), pp. 612–34.

Choi, D. and Hui, S.K. (2012). 'The role of surprise: understanding over- and underreactions using in-play soccer betting', HKUST/NYU Stern working paper.

Cohen, L. and Lou, D. (2012). 'Complicated firms', *Journal of Financial Economics*, vol. 104(2), pp. 383–400.

Corwin, S.A. and Coughenour, J.F. (2008). 'Limited attention and the allocation of effort in securities trading', *Journal of Finance*, vol. 63(6), pp. 3031–67.

Croxson, K. and Reade, J.J. (2011). 'Information and efficiency: goal arrival in soccer betting', ECONOMIC JOURNAL, forthcoming.

Da, Z., Engelberg, J. and Gao, P. (2011). 'In search of attention', *Journal of Finance*, vol. 66(5), pp. 1461–99.

DellaVigna, S. and Pollet, J.M. (2009). 'Investor inattention and Friday earnings announcements', *Journal of Finance*, vol. 64(2), pp. 709–49.

Garbade, K.D. and Silber, W.L. (1983). 'Price movements and price discovery in futures and cash markets', *Review of Economics and Statistics*, vol. 65(2), pp. 289–97.

Griffith, R.M. (1949). 'Odds adjustments by American horse-racing bettors', *American Journal of Psychology*, vol. 62(2), pp. 290–4.

Gromb, D. and Vayanos, D. (2010). 'Limits of arbitrage: the state of the theory', *Annual Review of Financial Economics*, vol. 2(2), pp. 251–75.

Harris, M. and Raviv, A.. (1993). 'Differences of opinion make a horse race', *Review of Financial Studies*, vol. 6 (3), pp. 473–506.

Hasbrouck, J. (1995). 'One security, many markets: determining the contributions to price discovery', *Journal of Finance*, vol. 50(4), pp. 1175–99.

Hirshleifer, D. and Teoh, S.H. (2003). 'Limited attention, information disclosure, and financial reporting', *Journal of Accounting and Economics*, vol. 36(1), pp. 337–86.

Hirshleifer, D., Lim, S.S. and Teoh, S.H. (2009). 'Driven to distraction: extraneous events and underreaction to earnings news', *Journal of Finance*, vol. 64(5), pp. 2289–325.

Hirshleifer, D., Lim, S.S. and Teoh, S.H. (2011). 'Limited investor attention and stock market misreactions to accounting information', *Review of Asset Pricing Studies*, vol. 1(1), pp. 35–73.

Hong, H. and Stein, J.C.(1999). 'A unified theory of underreaction, momentum trading, and overreaction in asset markets', *Journal of Finance*, vol. 54(6), pp. 2143–84.

Hong, H., Lim, T. and Stein, J.C. (2000). 'Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies', *Journal of Finance*, vol. 55(1), pp. 265–95.

Huang, L. and Liu, H. (2007). 'Rational inattention and portfolio selection', *Journal of Finance*, vol. 62(4), pp. 1999–2040.

Kahneman, D. (1973). *Attention and Effort*, Englewood Cliffs, NJ: Prentice Hall.

Kandel, E. and Pearson, N.D. (1995). 'Differential interpretations of public signals and trade in speculative markets', *Journal of Political Economy*, vol. 103(4), pp. 831–72.

Louis, H. and Sun, A.(2010). 'Investor inattention and market reaction to merger announcements', *Management Science*, vol. 56(10), pp. 1781–93.

Mondria, J. (2010). 'Portfolio choice, attention allocation, and price comovement', *Journal of Economic Theory*, vol. 145(1), pp. 1837–64.

Navon, D. and Miller, J.O. (1987). 'Role of outcome conflict in dual-task interference', *Journal of Experimental Psychology: Human Perception and Performance*, vol. 13(3), pp. 438–48.

Paschler, H. (1994). 'Dual-task interference in simple tasks: data and theory', *Psychological Bulletin*, vol. 116 (2), pp. 220–44.

Peng, L. and Xiong, W. (2006). 'Investor attention, overconfidence and category learning', *Journal of Financial Economics*, vol. 80(3), pp. 563–602.

Shin, H.S. (1991). 'Optimal betting odds against insider traders', ECONOMIC JOURNAL, vol. 101(408), pp. 1179–85.

Shin, H.S. (1992). 'Prices of state contingent claims with insider traders, and the favourite-longshot bias'. ECONOMIC JOURNAL, vol. 102(411), pp. 426–35.

Shin, H.S. (1993). 'Measuring the incidence of insider trading in a market for state-contingent claims', ECONOMIC JOURNAL, vol. 103(420), pp. 1141–53.

Simons, D.J. and Chabris, C.F. (1999). 'Gorillas in our midst: sustained inattentional blindness for dynamic events', *Perception*, vol. 28(9), pp. 1059–74.

Smith, M.A., Paton, D. and Vaughan Williams, L. (2006). 'Market efficiency in person-to-person betting', *Economica*, vol. 73(292), pp. 673–89.

Snowberg, E. and Wolfers, J. (2010). 'Explaining the favorite-long shot bias: is it risk love or misperceptions?', *Journal of Political Economy*, vol. 118(4), pp. 723–46.

Stein, J.C. (2009). 'Presidential address: sophisticated investors and market efficiency', *Journal of Finance*, vol. 64(4), pp. 1517–48.