UNFINISHED DRAFT: please treat as confidential; do not circulate or cite.

BBE: Simulating the Microstructural Market Dynamics of a Betting Exchange via Agent-Based Modelling

Dave Cliff
Department of Computer Science
University of Bristol
Bristol, BS8 1UB, UK
csdtc@bristol.ac.uk

***Abstract* — This paper describes the design and implementation of an agent-based simulation model of a contemporary online sports-betting exchange: such exchanges, closely related to the exchange mechanisms at the heart of major financial markets, have revolutionized the gambling industry in the past 20 years, but gathering sufficiently large quantities of rich and temporally high-resolution (sub-second interval) data from real exchanges – i.e., the sort of data that is needed in large quantities for contemporary machine learning techniques such as Deep Learning – is often very expensive, and sometimes simply impossible; this creates a need for plausibly realistic synthetic data, which is what the simulation described here is intended to provide. The simulator, named the *Bristol Betting Exchange* (BBE), is written in Python and the source-code has been freely released as an open-source project on GitHub: the intention is that BBE can become a common platform, a data-source and experimental test-bed, for any/all researchers studying the application of artificial intelligence (AI) and machine learning (ML) techniques to issues arising in betting exchanges. As far as I have been able to determine, BBE is the first of its kind: a free open-source agent-based simulation consisting of a model sports-betting exchange, a minimal simulation model of racetrack sporting events (e.g., horse-races or car-races) about which in-race bets may be made, and a population of simulated bettors who each form their own private evaluation of odds and place bets on the exchange before and – crucially – during the race itself (i.e., so-called "in-play" betting) and whose betting opinions change second-by-second as each race event unfolds. BBE is offered as a proof-of-concept simulator system that enables the generation of large high-resolution data-sets for automated discovery or improvement of profitable strategies for betting on sporting events via the application of AI/ML techniques: this paper closes with discussion of potential future work in this area. Illustrative data-sets generated by BBE have been made available, uploaded to Dryad at [DOI].**

Keywords— Betting Exchanges; Financial Markets; Automated Trading; Market Simulation.

I. Introduction

Like it or not, humans have been gambling for fun and profit (and sometimes ruin too, sadly) for a very long time: Hombas and Baloglou (2005) document the record of gambling in ancient Greek and ancient Egyptian cultures. Right now, more than four thousand years later, global revenues for the gambling industry are commonly estimated to be moving up in excess of US$500billion per year (PRNewswire 2020).

For the vast majority of the past four thousand years, gambling was a fairly simple and technology-free process: if two people *A* and *B* take opposing views on the outcome of some uncertain future event, such as the toss of an unbiased coin or the roll of a fair six-sided die, then they might agree a sum of money, the *wager* or *bet,* and then when the event's outcome is known either *A* will pay the bet to *B* or *B* pays *A*. Colloquially, we might say that *A* bet on one outcome (e.g. "*heads*", i.e., the tossed coin landing with the heads-side showing) while *B* bet on the complementary outcome ("*tails*"), but in the technical language that we'll use in this paper, we'll instead say that *A backed* heads (i.e., bet that heads *would* be the outcome) while B *layed* heads (i.e., bet that heads *would not* be the outcome).

If our two bettors *A* and *B* decide that instead of betting against each other they would prefer to take their chances with a commercial bookmaker (i.e., a "bookie"), either an individual or an organization, the same language applies: *A* could wager $100 with the bookie on a particular horse winning a race, or a particular team winning a soccer match, or any other uncertain event, in which case *A* is *backing* the specified outcome, while the bookie is *laying* the outcome. For very many years, this was most people's commonplace experience of commercial betting: a bettor backs an outcome with a bookie, taking the odds that the bookie specifies at the time, and the bookie lays the bet. In principle, it was possible for an individual bettor to make a lay bet: some bookmakers would offer a price (usually after being given time to work out the odds) on an outcome that is the complement-set of a single specific outcome – e.g. "I bet $100 that any team except for TeamX will win the Championship this year" – in such circumstances the bettor is laying Team X and the bookie is backing Team X, but for most people in everyday betting circumstances the usual implicit understanding is that bettors back and bookies lay.

Since the dot-com boom of the late 1990's, the worldwide betting industry has transformed from being focused almost entirely on traditional bookmaking as previously practiced for hundreds of years, to one in which the dominant practice revolves around bettors buying and selling bets on betting *exchanges*, and particularly on the internationally successful UK-based betting exchange *Betfair*, which is widely credited with being the disruptive innovator in this space, and which rapidly grew to huge financial success. The key innovation in Betfair was recognizing that the existence of a population of bettors with varying and opposing views, such that the various possible outcomes of a sporting event each attract some number of bettors willing to back that outcome, and some other number of bettors willing to lay the outcome (i.e. because they want to back some opposing outcome) is essentially the same situation as in a financial market where there are some number of traders interested in buying an asset, and some number of traders wanting to sell the asset. The reason why the services offered by Betfair and similar platforms are described as betting *exchanges* is because they bring backers and layers together to identify counterparties to a bet, and give all participants a shared summary view of the distribution of bets for a particular event, in a manner very similar to how most major financial exchanges bring together potential buyers and sellers, give them a summary view of the overall market supply and demand for some tradeable asset, and allows traders in the market to identify counterparties willing to transact at a price that both parties consider fair.

Briefly, almost all current electronic financial exchanges in the global markets operate what is known as a Limit Order Book (LOB) which aggregates and anonymizes the orders received from the various traders active in the market for a particular asset. Different traders will often have different prices at which they are willing to transact, and the quantities they have available to sell, or are willing to buy, will also differ: all of this is captured in the LOB, which is usually updated in real-time after each new order (or cancellation of a prior order) is received from a trader, with the updated LOB then immediately published to all traders simultaneously. In contemporary financial markets the LOB for a liquid asset will be changing and updated many times per second as traders issue new orders or cancel existing ones, but at any one instant the LOB displays a rich array of information showing all the prices at which traders in the market are willing to buy or sell, and the total quantities demanded or supplied at each of those prices: a trader looking at a snapshot of the LOB can see the state of the entire market at that time. Sometimes there is enough information in a single LOB snapshot for a trader to confidently make short-term predictions about which way transaction prices are about to move, e.g. because the LOB shows an imbalance between supply and demand in the market; but more generally it can be informative to view a sequence of changes in the LOB over some time period, i.e. as a time-series of snapshots, to identify and predict longer-term trends in supply, demand, and prices.

Betfair's key innovation was to create something very like a LOB, adapted to matching backers and layers in a betting market: at the heart of a betting exchange for a particular event is a data structure which is referred to as the *market* for that event, which is a direct analogue of the LOB in a financial market. For a track-racing event, the market will typically be displayed to a bettor as a rectangular grid of cells: with each competitor in the race allocated one row on the grid. A betting exchange's market for an event is split between backs and lays, arranged in order of goodness-of-odds, so that each cell in the grid displays a specific odds along with the total amount wagered at those odds: Figure BETLOB shows an illustrative example. Crucially, a betting exchange is not acting like a traditional bookie: it is not carrying risk of losing its own money by laying a customer's back, or backing a customer's lay: instead, it is acting merely as a centralized meeting and matching service for bettors to seek and identify willing counterparties with whom to bet.

INSERT FIGURE BETLOB HERE

Figure BETLOB: the "market" (i.e. the *order-book*) for a particular event on which bets are being made by bettors either backing or laying the event. The betting odds are shown as decimal values, the total return if the bet succeeds, explained further in Section II.

However an additional innovation made popular by Betfair, and the central focus of BBE, is so-called *in-play betting*, where customers can continue to trade on the exchange, laying or backing bets on the outcome of a sporting event such as a horse race or a football match after the event has started, right up until some previously-specified end-time such as the first horse passing the winning post or the final whistle in a soccer match. BBE has been specifically designed to model this in-play betting, and it is (as far as I can determine) unique in that regard: an issue discussed further in Section RELATEDWORK. One key aspect of in-play betting that BBE has been designed to explore is the *opinion dynamics* within the population of bettors; i.e., the extent to which the opinions of some bettors in the market for an event have their opinions (and hence their subsequent bets) affected by the distribution of money on the market for that event, and by any sudden changes in that distribution – because the distribution of money in the market for an event gives insight into the collective opinions, the overall sentiment, of the population of bettors active in that market. The relevant research literature on opinion dynamics is discussed further in Section RELATEDWORK.

From its foundation in June 2000, the Betfair exchange grew rapidly, to the point that by 2010 it was widely reported to be processing more transactions per day than all European stock exchanges combined (Roy, 2010), and various companies now offer technology tools and services for automated betting on the Betfair exchange, under the proposition that an automated system can be more profitable and less time-consuming than a human user manually placing bets and monitoring events.

For discussion of the growth and impact of betting exchanges such as Betfair see Davies, Pitt, Shapiro, & Watson (2005) and Cameron (2009). For a populist introduction to betting on Betfair, see Houghton (2006). The collections edited by Hausch, Lo, & Ziemba (1994), Vaughan Williams (2003, 2009), Hausch & Ziemba (2008), and Rodríguez, Humphreys, and Simmons (2017) provide detailed academic analyses of various aspects of these developments and related aspects of the sports betting industry: the review by Smith and Vaughan Williams (2008) is particularly relevant to the topics of this paper.

Devising profitable automated betting strategies is a labor-intensive activity requiring significant expertise in the design/development phase, and potentially needing access to very large amounts of Betfair exchange data, i.e. time-series of various betting markets on which strategies can be tested. Betfair does sell such data, but charges premium fees which can be prohibitive for non-professional betting-strategy developers, thereby erecting a major barrier to entry. A primary motivation for the design and development of BBE was to create a source of reliable synthetic data that could be used to explore the application and refinement of AI and ML methods to Betfair-style in-play betting-exchange datasets, thereby facilitating replacement of the skilled human betting-strategy designer with automated analysis, search and optimization processes: this is returned to in Section WHAT later in this paper. The motivation for BBE is not solely for saving money on data fees: for some data-intensive machine learning approaches, the amounts of data required simply exceed the amount that is plausibly available. For example, there might be some killer-app ML approach that will provably learn a hugely profitable bet-trading strategy, so long as you can provide it with the last 100 years of trading data from that betting exchange: because no exchange has been running for long enough, relying on real data from an exchange not a viable approach. If, instead, reliable synthetic data can be used to train the ML system to be a profitable automated bettor, that could then be tested/evaluated on the real data from the actual exchange, and the automated system then deployed to generate large pots of money

BBE involves three primary components: the minimally-simulated track-racing events that the bettors bet on; the betting exchange where backs and lays can be posted and matched; and the population of bettors themselves. These are discussed in Sections IV, V, and VI respectively. Section VII then presents illustrative results from the current BBE. Section VIII talks about future work that can build on the current BBE platform available on GitHub; and conclusions are drawn in Section IX. Before that, Section II gives further background information, and Section III reviews related work in the literature.

II. Background: Betting and Financial Markets

For completeness, we should first note here that an alternative way of organizing betting is as a *pari-mutuel* or *totalizer* system where the bookie pools all the money bet on a particular event such as a horse-race into a single "pot" of cash, deducts the bookie's fee from the pot, and then when the outcome of the event is known the money remaining in the pot is divided among those bettors who backed the correct outcome, with bettor payouts made pro-rata to the amounts wagered by each winning bettor, modulo the odds computed by the bookie working from the distribution of bets wagered. This style of betting is not modelled in BBE, because betting exchanges operate via a different mechanism.

It seems a fair assumption that readers of this paper do not require an explanation of the basics of how a track-race event operates, how it is set up and run: all of the aspects of a real track-race event that are captured in our model are described in precise detail in Section RACE-SIM.

The basics of betting on an exchange such as Betfair have already been introduced in the previous section, but the relationship between betting exchanges and financial exchanges does bear further discussion, to facilitate the review of related literature that follows in Section III. In the text that follows, I first give a brief illustrative description of the core operations at the heart of a contemporary electronic financial market, and then go on to discuss the issue of modelling the traders in a financial market, as a preamble to the issue of how best to model the bettors in a betting market.

The operation of a LOB-based financial exchange is best explained by an example. Consider the market for a fictitious asset with the unique identifying "ticker" code of XYZ, and say there are three sellers: seller S1 has issued an order to the exchange to sell 3 units of XYZ for an asking-price (the *ask*) of \$100 each; seller S2's order shows that she wants to sell 2 units for \$90; and seller S3 also wants to sell 2 units for \$90, so her order is the same as S2's, but happens to have been sent to the exchange after S2's was. The exchange's *matching engine* for XYZ would show the *ask-side* of the XYZ LOB as an ordered set of (price, quantity) pairs arranged in best-to-worst price order: ((£90,4), (\$100,3)) – note that the identity of the sellers is hidden, and that the orders from sellers S2 and S3 have been aggregated into a total quantity of 4 available at the single best-ask price of \$90. Continuing the same example, if there are four buyers in the market, B1 to B4, and their current orders (their *bids*) are 2 at \$75, 1 at \$80, 1 at \$75, and 3 at \$70, respectively, then the *bid-side* of the LOB would again aggregate and anonymize the orders and arrange them in best-to-worst price order, so the bid-side would be ((\$80,1), (\$75,3), (\$70,3)). If a Buyer B5 then came in with a bid-order of (\$100,5), the exchange's LOB matching engine would automatically match B5's bid with the asks received from S2 and S3, and with one of the units offered at \$100 by S1, and would remove those orders from the ask-side LOB (leaving it showing as ((\$100,2)) – the residual of S1's ask) and would also notify the traders concerned that their orders had been matched with a counterparty, at which point their transactions go into a "clearing" process that handles the necessary transfers of ownership and of funds. If instead B5's order had been a bid of (\$100,3), the matching engine would have consumed all of S2's two items offered at \$90 and one of the two offered at the same price by S3; in this case S2's order is matched before S3's, because S2's order arrived at the exchange before S3's: that is, the exchange's engine is here operating a *time-priority* matching process. For more detailed discussion of LOB dynamics, see Gould *et al.* (2013), Abergel *et al.* (2016) and Bouchaud *et al.* (2018)

The market for an event in a betting exchange such as Betfair is a close correlate of the LOB: instead of traders issuing orders to the financial exchange that each signal a desired direction (buy/sell), price, and quantity, bettors on a betting exchange issue orders that each state a direction (lay/back), the odds (also frequently referred to as the *price*), and the *stake* (i.e., a quantity of currency).

In this paper, as on most major betting exchanges, all odds will be expressed as decimals (potential total returned), rather than using fractional or American representations. For example, where a successful bet with a \$1 stake generates winnings of \$10 plus the original stake returned, for a total of \$11, the decimal representation of the odds is 11 (fractional: 10/1; American: +1000); similarly, where a successful bet with a \$5 stake returns winnings of \$1 plus the original stake returned for a total of \$6, the decimal odds are 1.2 (fractional: 1/5; American: -500).

For any one sports-betting market, the betting exchange will have a matching engine running continuously that receives orders from the bettors, and aggregates and anonymizes them in a way directly analogous to the LOB matching-engine in a financial exchange: rather than show for each price the total quantities offered and bid-for by sellers and buyers respectively, the betting-exchange market shows, for each odds-price backed or laid, the total amount of currency staked at that price; and, as in a financial market, when it comes to matching orders from one counterparty to another, time-priority matching is commonly used. Further details of the implementation of the BBE matching engine are given in Section MATCH-ENG; before that, in Section RELATED, we review the literature and public-domain software for financial-exchange simulations, which have both proven to be useful resources in the construction of BBE.

Just as one would expect an agent-based model of a financial exchange to include not only a simulation of an exchange, but also some number of simulated traders issuing bid and ask orders to the exchange, thereby populating its LOB, and eventually to enter into trades with one another; so in BBE we need not only a simulation of the betting exchange's matching engine to give a continuously updating order book for that race-event's market, but also some number of simulated bettors to each form an opinion on the likely outcome of the event, and to place back and lay bets accordingly.

As is discussed in more detail in the next section, while there is a rich seam of literature on simulating traders for financial markets, there seems to be almost no literature on simulating bettors for betting markets. We will review the literature in more depth in the next section, but before that it is useful to explicitly tease out the extent to which a group of bettors placing bets on a betting exchange can be viewed as a system in which *opinion dynamics* is a key consideration. There is a long-established literature on opinion dynamics, but it tends to involve simple abstract models typically involving some number of agents (the *population*) interacting in such a way that the opinion of Agent A may to some extent influence the opinion of Agent B, depending on specific circumstances, and possibly B would at the same time exert some influence on the opinion of A, and as the model evolves through time some sequence of these pairwise opinion-altering interactions happens in the model system such that the overall distribution of the population changes in interesting ways over the duration of the simulation experiment. Much of this opinion dynamics work has been directed at identifying the conditions under which consensus can be reliably reached (e.g., DeGroot, 1974) and the conditions under which extreme opinions might form and spread within a population (e.g., Deffuant, 2006; Meadows & Cliff, 2012, 2013).

Opinion dynamics modelling techniques have only very recently been integrated into models of financial markets: (see e.g. Lomas & Cliff, 2021) which was motivated by the recent groundbreaking work of Nobel-Prize-winner Robert Shiller (2017, 2019) on *Narrative Economics*, the attempt to better understand economically unfathomable behaviors (such as investing in an intrinsically valueless 'asset' such as Bitcoin) as a function not of rational economic reasoning but rather as a function of the narratives, the stories, that economic agents tell each other and themselves in justifying their buying or selling decisions: these narratives are nothing more than external verbalizations of internally held opinions, so Shiller's work on narrative economics is manifestly opinion dynamics by another name.

Yet, to the best of my knowledge, the communities of researchers involved in agent-based modelling of opinion dynamics, and in the nascent field of narrative economics, have both thus far overlooked betting markets as a subject of study: this seems to me to be something of an oversight: surely a back or a lay is nothing more than an opinion expressed in the most concrete terms, a statement of how much money the bettor is willing to risk losing if their opinion is wrong? Surely a betting exchange, the market for an event, is just a set of monetised opinions, is not much more than the consequence of a bunch of bettors each willing to put their money where their mouth is?

For this reason it seems reasonable to propose here that agent-based models of betting exchanges, such as BBE, should be considered first-class instances of systems in which opinion dynamics can usefully be studied: all that is needed is for the population of simulated bettors to interact with one another in ways similar or identical to those already studied in the opinion dynamics literature: this would be a model of individual bettors interacting with one another, sharing tips on which competitor might win and which might lose, in a peer-to-peer fashion. That would be one aspect of BBE that closely matches traditional opinion dynamics studies, but the centralised order-book of the betting exchange, the market for a specific event, is a second form of opinion-influencer that is not traditionally studied in the opinion dynamics literature: surely if each bettor in the can view the array of backs and lays that make up the market at any one instant in time, and if each bettor can also observe changes in the market over time, then it is plausible that the instantaneous distribution of bets, or any changes in that distribution, could or should have an influence on the opinions of at least some bettors?

To give an extreme example: bettor *B* may start out quite sure of her opinion that horse *H1* will win the race, and she backs it accordingly; but if partway through the in-play betting on this race the amount backing *H1* stays constant but the amount wagered in lays against H1 suddenly goes up sharply, then (human nature being what it is) there is a fair chance that *B's* opinion of which horse is most likely to win will change, and she might hurriedly make a fresh bet to try to compensate for this. Here the distribution of stakes in the market is a reflection of the bettor-population's overall collective *sentiment* about the outcome of the event, and that representation of the sentiment in turn can alter the opinion of some or all bettors in the market: this feedback loop could be positive or negative for any one bettor, depending on that bettor's dispositional characteristics such as her certainty or confidence in her own opinion. Analysis of collective sentiment has been an active topic of research in financial markets for many years (see e.g. Mitra & Yu, 2016, Pozzi *et al.*, 2017; Birbeck *et al.,* 2018; Liu, 2020), but sentiment analysis of bettor opinions from time-series of betting-exchange order-books seems to be wholly less researched.

Furthermore, with only the slightest change of language, we could say that when a bettor *B1* backs some outcome *O*, she is *predicting* that *O will* happen; and similarly that when a bettor *B2* lays *O,* so *B2* is *predicting* that *O* will *not* happen. In this characterization, there is no requirement for *O* to be a sporting event; it could as easily be the outcome of a political election, or which movie will win this year's best-picture Oscar award, or what the annual profits of a particular company will be when they are next reported, or what that company's share-price will be three months from now. This exposes the natural link between betting exchanges and online prediction markets such as the long-established *Iowa Electronic Markets* (2021), and more recent entrants such as *PredictIt* (2021) and *Smarkets* (2021): for a review of research issues of prediction markets in corporate settings, see O'Leary (2012). Technically, prediction market usually allow the participants to trade in a *contingent security*, i.e. to buy or sell units of a tradeable asset that will have a specified nonzero value if the specified outcome does occur (that is, the end-payout is contingent on the outcome) and will have a zero value for all other outcomes, and hence any such prediction market could also be argued to represent a form of parimutuel betting system. Moreover, the fact that the outcome *O* could be a *financial* outcome, concerning the price of an asset or the balance-sheet of a company, brings us back to the world of finance, in which a simple contingent security is more commonly known as a *binary option* (binary because it pays either something or nothing, depending on some specific outcome condition): most exchange-traded derivative contracts such as binary options are, put bluntly, just glorified gambling – even if the participants in those markets like to think differently.

III. Related Work

I have searched online[1] to the best of my ability for academic papers describing something similar to BBE, but have found no good matches. It is the case that there are various simulators offered by computer-gaming companies and by betting-exchange operators which allow a bettor to bet on simulated sporting events, using simulated money, either for pure fun or to practice betting without risking loss of real money, and several bookmakers also offer real-money betting on "virtual" sporting events where the competitors are synthetic within a simulation or e-sports computer-game, but these such simulators do not offer the possibility of modelling each bettor's changing opinions and issuing of new bets during an in-play betting session.

BBE's simulation of a track-racing event is a deliberately minimal mathematical/algorithmic model, intended purely as bare-minimum mechanism capable of generating data that are plausibly similar to those that would be expected from a real track-racing event. It is described in detail in Section WHAT, and the Python source-code is available at (BBE, 2021), but it is not in itself intended as a contribution to the simulation literature. As was mentioned in the previous section, there are various simulators made available to bettors for pure fun, for practice-evaluation of betting strategies intended for subsequent deployment on real race-betting, and for real-money betting on computer-synthesized racing events presented to the bettors via CGI-rendered movies of the simulated races. To be of value to the user, the practice and for-fun systems involve some kind of simulation of betting on an exchange, at least from the user-interface perspective, whereas the real-money betting on synthetic CGI sporting events involve actual links to a real-world betting exchange, but these are all proprietary systems: their source-code is not open-sourced, and they are not offered as (and typically cannot be used as) generators of synthetic data for use in subsequent AI/ML-based research on automated betting strategies. Despite repeated and extensive online searches, I have found no reports of a sports-betting simulator that is even tangentially comparable to BBE.

Despite this, there is some solace to be found in the literature on simulated *financial* exchanges: given the manifest similarity between betting exchanges and financial exchanges, the literature on financial-market simulation is of some relevance, and indeed in that field there is a tradition of researchers freely releasing public-domain open-source simulators as common platforms for research. Accurate simulators for the kind of LOB-based financial exchange explained above are found in the publications and associated open source-code releases for the simulators *OpEx* (De Luca *et al.,* 2011; De Luca & Cliff, 2011a, 2011b), *BSE* (BSE, 2012; Cliff, 2018), and *ExPo* (e.g., Stotter *et al.,* 2014), which each aimed to make public and free the kind of financial-exchange simulator that was previously private and proprietary when used in experiments conducted at major corporate research lab such as IBM (e.g., Das *et al.*, 2001) and Hewlett-Packard (e.g., Cliff & Preist, 1998). The BBE betting-exchange simulator described in detail in the remainder of this paper has benefitted from the lessons learned from (and the open-source code releases of) OpEx, ExPo, and BSE.

The similarities between financial markets and betting markets, and factors of interest that are common to the two types of markets, have been studied by academic economists, statisticians, pollsters, and forecasters: see, for example (Piron & Smith, 1993; Hurley & McDonough, 1996; Thorp, 2006; Camerer, 2008; Tompkins *et al.*, 2008; Palomino *et al.*, 2009; Franck *et al.,* 2010; Smith & Vaughan Williams, 2010; Haahr, 2011; Brown *et al.,* 2016; Flepp *et al.*, 2016; Sung, *et al.,* 2016; Wall *et al*., 2017; Hanke *et al.,* 2018; Restocchi, 2018; Sung *et al.,* 2018; Elaad *et al.*, 2020). Although all of these studies involved analysis of data from real-world betting exchanges (or other gambling mechanisms) rather than working with synthetic data from a simulator. When such academic work does talk of simulation studies, this almost always involves the use of historical real-world time-series data as input to a process that then simulates a betting strategy operating on that real-world data over a number of days or weeks; I know of no such studies that simulate some number of heterogeneous bettors each with their own strategies for in-play betting.

Possibly one reason for this apparent lack of agent-based models of betting-exchange systems is the lack of any literature on individual bettor-agent models or strategies. This is in stark contrast to the literature on autonomous trader-agents for financial markets where there is a 30-year tradition of publishing details of trader-agent strategies, each intended in one way or another as an improvement on the previous best. A tradition has emerged in this body of literature of giving each different strategy an abbreviated ticker-code style of name; the most notable strategies in the public domain are: *Sniper* (Rust *et al.*, 1992); *ZIC* (Gode & Sunder, 1993); *ZIP* (Cliff, 1997); *GD* (Gjerstad & Dickhaut, 1997); *MGD* (Tesauro & Das, 2001); *GDX* (Tesauro & Bredin, 2002); *HBL* (Gjerstad, 2003); AA (Vytelingum *et al.*, 2008); *RE* (e.g., Pentapalli, 2008); *SHVR* and *GVWY* (BSE, 2012; Cliff, 2018); *ASAD* (Stotter *et al.*, 2014), and *PRZI* (Cliff, 2021). All of these trader-agent models were

[1] Have a footnote here that explains where was searched (arxiv, google scholar), what terms were used, and what the responses were.

essentially hand-designed, although from the mid-1990s onwards it became commonplace to incorporate adaptation mechanisms drawn from machine learning or statistical analytics, so that each individual trader could adapt to the particular market circumstances it found itself embedded in, and so that traders could respond to changes in market conditions as they happened. Much of this work on artificial trader-agents (now commonly referred to as *robot traders*) was intellectually underwritten by the large body of prior work in *experimental economics*, in which controlled repeatable empirical studies teased out the fine details of human trading behaviors in realistic laboratory models of financial markets (see, e.g.: Smith, 1962, 2000; Kagel & Roth, 1997; Plott & Smith, 2008; and Hommes & LeBaron, 2018).

In comparison, for betting markets, a substantial body of research has been published in various fields that reports on empirical studies of the behavior of actual human bettors (see, e.g., Kanto *et al.*, 1992; Swidler & Shaw, 1995; Bradley, 2003; Julien & Salonie, 2008; Choi & Hui, 2014; Feess *et al.*, 2015; Brown & Yang, 2016; Suhonen *et al.,* 2018; Merz *et al.,* 2020), where there is a common concept of *the representative bettor*, i.e. an idealisation of the betting behavior of the average bettor, and in which *prospect theory* (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) has been a major influence. In this area of the literature, a fair amount of effort has been expended on exploring and explaining *the favourite-longshot bias*, where bettors tend to undervalue favourites (outcomes that have short odds, high probability) and overvalue outsiders (outcomes that have long odds, or low probability), a bias that is seen repeatedly in betting markets and in financial markets, but there is not a long-established literature that is comparable to the body of work on robot traders for financial markets, where researchers developed the succession of specific automated trading strategies named in the preceding paragraph.

In recent years there has been a growing body of research publications exploring the use of statistical approaches, machine learning, and/or artificial intelligence, in betting markets. Various authors have reported mathematical or algorithmic approaches to profitable betting or trading on betting exchanges, often involving machine learning; see, e.g.: (Easton & Uylangco, 2009; Ioulianou *et al.,* 2011; Aruajo-Santos, 2014; Tsimpras, 2015; Dzalbs & Kalganova, 2018; Bunker & Susnjak, 2019; Hubáček *et al.*, 2019; Axen & Corris, 2020; Goncalves *et al.*, 2020; Hubáček & Šír, 2020; Wheatcroft, 2020; and Wilkens, 2020). However all of these studies work from databases of historical odds/price time-series from one or more betting exchanges, and none of them report on methods for in-play betting: in this sense, they are comparable to automated methods for identifying buy and sell signals from analysis of historic time-series of daily price-movements in financial markets. CHECK THIS.

Such an approach is perfectly valid, of course, but it gives little or no insight on how best to trade second-by-second in a fast-moving situation such as a market for a heavily-traded asset, or an in-play betting market for an event that is underway. Furthermore, such an approach fails to capture the close-coupled feedback loop where events occurring mid-race cause some bettors to shift their positions, placing new in-play bets, which are then visible to other bettors in the market, causing those other bettors to also adjust their positions. And, fundamentally, even the highest-resolution time-series of in-play betting prices for a specific event is only half the story: without a similarly accurate record of how the event itself played out (e.g., second-by-second records of the positions of the competitors on the track), there is simply nothing to correlate the betting activity against. For this, then, we need a simulator that (re-)creates the kind of events that bettors like to bet on. Although in principle we could later extend BBE to incorporate simulations of sporting events such as soccer games or tennis matches, we started with the track-race events because they are relatively straightforward to characterise mathematically, as described in the next section.

IV. The Race Simulator

As currently configured, BBE is an abstract minimal model of some number of bettors interacting via a betting exchange to back and lay bets on the outcomes of racing events. The model is sufficiently minimal and abstract that in principle it could be interpreted as a model of gambling on horse-races or greyhound-racing, two sports in which betting is deeply embedded; or it could equally be interpreted as a model of gambling on races between vehicles such as in Formula-1 or NASCAR car-racing, motorbike racing, cycling races; or human track-and-field athletics running-races; or any other type of event where some number of participants are started at the same time and then compete to cross a finish-line first. There is nothing in our model that specifically limits us to one specific type of race, so we will talk here just of races and competitors.

Let $\mathbf{C}$ be the pool of competitors, and let the number of competitors in $\mathbf{C}$ be $C=|\boldsymbol{C}|$. For any one race, denoted by subscript r, some number n_r of competitors are drawn from $\mathbf{C}$ and then those individuals compete by racing along a one-dimensional track of specified length L_r: the position along the track at time t of competitor c is a real-valued distance denoted by $d_c(t)$, and the state of the race at time t can be summarized by the vector $\boldsymbol{d}(t)$ in which the i^{th} element is $d_i(t)$ and hence $|\boldsymbol{d}(t)|=n_r$. Individual competitors are merely represented as points along the track: they have no physical extent in our model, although they can impede or block one another's progress, as described further below.

A competitive race starts at time t=0, and the clock then continues to run until the last competitor c achieves a position $d_c(t) >= L_r$ – i.e., the race ends when the slowest competitor crosses the finish line; in-play betting may be specified to end at that at time, or possibly when an earlier condition is met, such as the third-placed competitor passing the line. The initial value of $\boldsymbol{d}(t)$ might be a zero vector (i.e., all competitors lined up on the start-line, or all horses stood in a starting-gate) or not (e.g. some motor-races start with competitors spread out spatially over a "grid" with the better competitors to the front; and some horse-races start with the competitors gathered stationary behind a tape "barrier", at various distances from the tape).

Each competitor's progress within a race is governed by a discrete-time process such that $d_c(t+1)= d_c(t)+S_c(t)$ where $S_c(t)$ is a function that generates a forward-step for competitor c at time t: $S_c(t)>0$ at all times, to ensure that the race will eventually end, and should usually be a stochastic function so that the outcome of the race cannot be determined precisely at the start. For example, using $\boldsymbol{U}(min,max)$ to represent a uniformly distributed random variable over the range [*min*, *max*], competitor *c1* might have $S_{c1}(t)=\boldsymbol{U}(10,20)$ while competitor *c2* might instead have $S_{c2}(t)=\boldsymbol{U}(1,25)$: given the specifications of these two *S* functions, we can say that one competitor is more or less likely to cross the finish line first on the average, but we cannot say for sure who will win a specific individual race.

In real races, any one competitor will be more or less suited to the details of a specific race: some competitors will do better in shorter races, while others will prefer longer distances; some will prefer flat courses, others will prefer undulations or hills; some will prefer racing in warm weather, others would prefer it to be cool; and so on. For studies of distance and pace preferences in horse-racing, see Benter, Miel, & Turbough (1996) and Edelman (2008, 2009), and for analysis of how racetrack surface conditions affect racehorse performance see Maeda *et al.* (2012). To model these effects, our simulator allows some number *f* of performance-factors that modulate the $S_c(t)$ function for a competitor: in any one race *r*, the *f* factors will each take on specific values (i.e., a particular length, a particular flatness, a particular temperature, etc) but these are all abstracted and normalized to simply be a vector $\boldsymbol{f}_r$ of length *f*, where each value in the vector lies in the range [0,1], and each competitor's *S* function is extended to incorporate $\boldsymbol{f}_r$, s.t. $d_c(t+1)=d_c(t)+S_c(t,\boldsymbol{f}_r)$. One simple illustrative instantiation of how $\boldsymbol{f}$ can be used within the *S* function is for each competitor *c* to have a *f*-dimensional *preference vector* $\boldsymbol{p}_c$ specified when the competitor is created, and for the step-size to be modulated by the Cartesian distance between $\boldsymbol{p}_c$ and $\boldsymbol{f}_r$, e.g. $S_c(t,\boldsymbol{f}_r)=(k-|\boldsymbol{f}_r-\boldsymbol{p}_c|)\boldsymbol{U}(d_{min},d_{max})$ with the constant $k>0$ – i.e., the farther the current race's factor-vector is away from the individual the individual competitor's preference vector, the smaller that competitor's step-sizes will be in this race.

For the races to present interesting betting opportunities we need the abilities of the various competitors to be sufficiently closely matched (given the race conditions) that the outcome is not obvious *a priori*. In horse-racing, jockeys wear weights as handicaps, and many races are restricted to particular types of horses (e.g. "female horses no older than three years who have not won more than £500 in the past"). These measures are intended to approximately balance the field, so that it is not absolutely obvious at the outset which runner will win. Our simulations are similarly constrained, otherwise we couldn't realistically expect our simulated bettors to interact with one another very much.

Because we are interested in simulating in-play betting, we also add a within-race dynamic to each competitor's performance. This allows us to model, for example, situations in which a competitor takes some time to build up speed to maximum pace, and/or grows tired toward the end of the race (after an initial plateau, performance starts to gradually reduce) and/or makes a final burst of effort when in sight of the finish line (i.e., performance is briefly increased, until the end of the race). We model this by introducing for each competitor a time-varying *responsiveness* factor $r_c(t)$ which modulates $S_c(t,\boldsymbol{f}_r)$. As just described, responsiveness might vary endogenously (i.e., due to factors internal to the competitor) but we are also interested in situations where responsiveness is affected by exogenous factors, i.e. due to causes external to the competitor: in the extreme, an exogenous factor could reduce a competitor's responsiveness to zero, for example if a horse refuses or falls at a fence in jump-racing, or if one race-car is crashed into by another competitor at a tricky corner. For this reason our responsiveness function r_c for competitor *c* should also take into account not only *c*'s current position but also the positions of all other competitors in the race, i.e. the full vector $\boldsymbol{d}(t)$, and hence we have:

$$S_c(t,\boldsymbol{f}_r,\boldsymbol{d})= r_c(t,\boldsymbol{d}).(k-|\boldsymbol{f}_r-\boldsymbol{p}_c|)\boldsymbol{U}(d_{min},d_{max})$$

In most circumstances $r_c(t)\sim 1.0$ but a competitor's responsiveness can also be set to zero to model the situations where a competitor suddenly stops mid-race because of injury or mechanical failure.

As described thus far, our simulated races are similar to short-distance sprints on an athletics track, where each runner is assigned a lane to stay in, and hence there is an expectation that there will be no physical interactions between the n_r competitors in the race: and, given the lack of any interaction between the competitors, the winner could in principle be decided by a sequence of n_r single-competitor time-trials. However in longer-distance human athletics races, and in pretty-much all animal or vehicle races, the runners are not constrained to individual lanes and hence they can interact with each other on the track to create unexpected outcomes: the favourite at the start of the race might find themselves boxed-in by other competitors (this routinely happens in animal racing and in vehicle-racing too), or delayed trying to find an opportunity to overtake/lap a slower competitor (common in car races). Our model of in-play betting would be made richer by introducing some kind of in-race competitor interactions on the track. As a first approximation, we have introduced a mechanism where at each time-step for each competitor a value $g_c(t)$ is computed, representing "ground lost due to delays by other competitors" such that $g_c(t)$ is a coefficient in [0,1] which modulates the competitor's track-distance delta, its step-forward this timestep, such that $g_c(t)$=1 means no delay at all and $g_c(t)$=0 is the opposite extreme, no forward movement at all on this timestep because of interference with competitors. Let *P1* denote a competitor that is leading a race, *P2* denote the competitor that is currently in second-place, and more generally let *Pn* denote the *n*-th place competitor. Then to model obstruction and boxing-in we can employ some attenuation function such that a competitor's value of $g_c(t)$ lowers in proportion to the proximity of other competitors in front of that competitor – so a *P1* competitor isn't interfered with at all; a competitor *P2* with "clean air" in front of it (i.e., a long distance behind *P1*) is similarly unhindered; but a *P2* immediately behind the *P1* (or, more generally, any runner in position *Pn* that has one or more runners at positions *P(n-1)* or less in front of it and nearby) has a much higher chance of $g_c(t)$ being lower than 1, to model the *P1* runner interfering with the progress of *P2*. This needs to be probabilistic (it's not *guaranteed* that a runner is impeded by the runner in front) and should take into account the *number* of runners in close proximity, as well as their distances. For that reason, and also because some types of race require the competitors to each start in a designated lane/position and possibly also to remain in-lane until some condition is met (e.g. in Olympic 800m track-running, the competitors each start in individually-assigned lanes and must remain in them until they exit the first turn of the track), the $g_c(t)$ function will also need the current track-positions of the competitors, i.e. the vector $\boldsymbol{d}(t)$, And, finally, to be maximally general we will replace the $\boldsymbol{U}(d_{min},d_{max})$ uniform-distributed random step-size component with a generic step-size function $\boldsymbol{\Delta}$ that takes some vector of parameters $\boldsymbol{v}_{\Delta}$, and hence our step-forward distance-increment function for competitor *c* is:

$$S_c(t,\boldsymbol{f}_r,\boldsymbol{d}) = g_c(t,\boldsymbol{d}).r_c(t,\boldsymbol{d}).(k-|\boldsymbol{f}_r-\boldsymbol{p}_c|)\ \boldsymbol{\Delta}\ (\boldsymbol{v}_{\Delta})$$

Details of the functions $g_c(t,\boldsymbol{d})$ and $r_c(t)$ used in our simulations are given later in this paper; and although all our work thus far has used $\boldsymbol{v}_{\Delta} = (d_{min}, d_{max})^T$ and $\Delta = \boldsymbol{U}(\boldsymbol{v}_{\Delta}{}^T)$, we note here that one avenue for further work is to explore other distributions for the stochastic step-size.

While it would be desirable to derive clean closed-form equations for the probability of any specific competitor winning any particular race, this is simply not practicable because plausible functions for $g_c(t)$ and $r_c(t)$ will be nonlinear, and the competitors are deliberately closely matched and hence are very likely to interact in nonlinear ways within any one race. Instead, we can empirically determine reliable estimates of the true underlying probabilities by running a sufficiently large number *N* of repeated i.i.d. "practice" simulations of a specific race with a specific set of competitors, and then calculating summary statistics from the data thus generated: in the simplest case, if after *N* i.i.d. practice simulations of the same race a competitor *c* has won the race in *w* of those simulations, then our estimate of the probability of *c* winning this race in the one "competitive" simulation in which actual betting occurs is *w/N*, and this estimate could be used by a better to set the starting price (i.e. the *ex ante* odds) for that competitor. Performing the necessary number of i.i.d. practice simulations is computationally expensive on a single CPU -- it will run slow, potentially slower than real time, but it’s embarrassingly parallelizable so it could be done faster by spreading the work over multiple machines. Alternatively , a single ensemble of *N* i.i.d. practice simulations could be run to establish the "ground truth" for the race at this point in time, and then each bettor's estimate of the outcome could be formed by injecting bettor-specific noise and perturbation factors to distort the bettor's estimate away from the ground truth. If the noise and perturbation factors are heterogeneously distributed across the population of bettors, this would have broadly the same effect in the simulator for less computational cost.

Just to be clear, there is no claim being made here that real human bettors run multiple i.i.d. simulations of a race in their head before they decide on the likelihood of a particular competitor winning a specific race, nor that they have the absolute ground truth magically revealed to them, which they then screw up: humans' evaluation of probabilities and risk, and the way in which humans form and change their opinions about possible future events, has been the subject of much study and the mechanisms described here are definitely not intended as a contribution to those studies. Rather, what is described here is workably simple algorithmic method for arriving at the same situation as we see in a betting market: i.e., some number of bettors each with their own privately-held internal opinion about the likely winner of a race.

Manifestly, much of the dynamics of a betting exchange is driven by differences of opinion among the bettors. So long as there is sufficient inherent uncertainty in the outcome of a simulated race, the differences in opinion can be engendered in various ways: one method would be to give each bettor *b* only a small number N_b of i.i.d. practice-runs of the race in which to form its opinion of the odds of success for each competitor – i.e. for the individual sample-size to be sufficiently small that the bettor's estimates vary significantly; another method would be for each bettor *b* to maintain their own private internal estimates of the values of the various constituents of the function S_c (such as the preference vector $\boldsymbol{p}_c$) for each competitor in the race, where *b*'s estimates differ from the true values, such that when *b*'s N_b i.i.d. practice-runs generate its estimates of the probability of winning for each of the competitors, the resultant probabilities differ from the true underlying probabilities because of *b*'s inaccurate estimates – this would model a bettor who is consistently over-optimistic or consistently over-pessimistic about a particular competitor's chances in a particular type of race.

Of course, for in-play betting, the bettors will need also to revise their estimates of the probable outcome of the race for each competitor while the race is underway. In principle, this could be done on every time-step, although in practice it can save computational effort if each bettor revises its opinions less frequently than that. This can be done using the same method as was used to create the bettors' ex ante opinions about likely outcomes: each bettor runs N_b practice-races to form an estimate of the probabilities of winning for each runner, *given where they currently are on the track* – so if a competitor A that started as the favourite (i.e., most likely to win) happens to have been delayed by other runners and is now in the mid-field, and the second-favourite competitor B has pulled into a clear lead, the odds for A winning should be lengthened and the odds for B winning should be shortened. Each bettor compares its revised estimates of the odds with those that are on the betting exchange and maybe revises/cancels/adds bets to the betting-exchange order-book accordingly.

As described thus far, we have a way of creating some number of races where each race involves some number of competitors from **C**: this is enough to generate a sequence of sporting events on which bettors can gamble, but we need also to have a similarly minimal abstract model of a population of bettors **B** who gamble on these events. Just as not every competitor in **C** will be involved in each race, so there is no need for every bettor to gamble on each event: for any one event, some number of bettors can be drawn from **B**. Fundamentally, a betting exchange operates by bringing together bettors with opposing views or beliefs: if bettor *b1* believes that competitor *c1* will win the current race, and bettor *b2* believes that *c1* will not win the race, then *b1* would *back* (i.e., bet-to-win) *c1* while *b2* would *lay* (i.e., bet-to-lose) *c1*. The exchange can then match *b1*'s back with *b2*'s lay, and take a small commission for doing so: both *b1* and *b2* have found a willing counterparty to take the other side of their bet, so they are each happy customers.

In our simulation, each bettor *b* in B makes predictions about the outcome of a race and bets on the basis of those predictions. Intuitively, the accuracy of an individual bettor's predictions can be situated on a continuum from making equiprobable random choices over the space of possible outcomes for a particular race (thereby totally ignoring all available information about the nature of the race and about each of the competitors) through to a god-like omniscient bettor who has perfect information on all factors that contribute to the outcome of the race (which, in our abstract model, would be full and precise details of the race's factor vector $\boldsymbol{f}$, each competitor's preference vector $\boldsymbol{p}$, and the details of each competitors step-function S). One way of distributing the population of bettors along this continuum is to initially make each bettor form equiprobable estimates of the likelihood of each outcome for a race, and then to randomly allocate each bettor some number d of "dry-run" trials: in any one dry-run, the race is simulated and that bettor use the outcome of that simulated race to revise its estimate of the outcome when the race actually takes place. A bettor with d=0 remains a purely random bettor; a bettor with d=1 has one trial's worth of data to go on, which is better than nothing but is not as good as d=10 or d=100; in the limit, as d approaches infinity, the trial-outcome information that is available to an individual bettor is so extensive that accurate estimates of the probability of each possible outcome for the race can readily be made.

Each bettor is endowed with some initial funds to be used in making bets, and bettors then submit back or lay "orders" (bets) to the exchange for each race, and the exchange's internal matching engine duly arranges the array of orders received, matching backs and lays as appropriate, and after the race is over the exchange takes care of the necessary book-keeping, updating the balances of the various bettors to account for their wins and losses.

As explained thus far, the simulator can be seen to consist of the betting exchange, the race simulator with its pool of competitors and races, and the population of bettors with their varying degrees of accuracy of estimating the likelihood of different outcomes.

In comparison to a financial market, the uncertain evolution of the positions of the competitors in the race as it unfolds is analogous to the moment-by-moment movements in the price of some tradeable financial asset; and the internal mechanisms of the betting exchange implement functions largely identical to those performed by the matching engine in a financial exchange: i.e., receiving orders from traders and aggregating them at different price-points and anonymizing them and then arranging them into a public/published limit order book (LOB) is much the same as receiving bets from bettors and aggregating, anonymizing, and arranging them onto the "book" showing the "market" for bets on a particular race.

Thus far we've described the component within BBE that is analogous to the dynamically-varying asset-prices in a financial market. Next, we'll look briefly at the model of the central exchange platform that collects orders and aggregates them into the order-book; and after that we discuss the largely uncharted waters of modelling bettors for in-play markets.

V. The Betting Exchange

Blah blah blah: cite BSE and other financial exchange simulators. Explain the differences. Keep this section brief -- efer to the GitHub code for full disclosure of how it works,

VI. Modelling Bettors

Note that there is a bunch of different public-domain trading-agent strategies for CDA financial markets, but there is no similar body of work/literature for betting-agent strategies: we're kind of on our own here; anything we do will be a first step in the right direction, and ideally BBE will ignite/enable work where researchers compete to create ever-better robot-bettors, comparable to ZIC/ZIP/GD/GDX/etc. We need as many of these as we can get, but we can also use the "representative bettor" work mentioned in Section WHAT, and we could also add in "noise traders" (Compare here to Ash B + Franck McG paper on Ash's thesis simulator, and also Frank's PhD student on stylised facts of betting markets).

One point of difference with financial-trading agents is that our bettors are trying to make money not by getting profit on executing a sales-trader assignment with a predefined limit-price but instead they're trying to get a payoff from the bets they hold when the race ends – this is a market for a contingent security.

VII. Illustrative Results

Blah blah blah some sample results – show spacetime plots of an illustrative selection of races, then the to-win odds from N=10, N=100, and N=1000 runs to illustrate convergence to ground truth. Maybe also place odds?

Show PRZI function used for responsiveness curves.

Show illustrative bettor beliefs/opinions over duration of race, for a race where a back-marker comes through to win, or where the leader suddenly fades and loses, to show dynamic moves in their beliefs as events unfold in the race – od this for several heterogeneous bettors to show variation.

Validation/verification:s xref stylized facts?.

VIII. Further Work

STRESS THIS PAPER IS JUST ABOUT THE DESIGN & IMPLEMENTATION OF THE SIMULATOR, to establish the background: DETAILS OF THE RESULTS FROM SPECIFIC USE-CASES AND/OR DIFFERENT IMPLEMENTATIONS/ EXTENSIONS OF BBE WILL COME IN LATER PAPERS (e.g. with MEng project students).

BBE is sufficient to generate large amounts of data that can be used in machine learning. The next thing to be addressed is exactly what sort of machine learning we are going to use.

REST OF THIS TEXT COMES FROM OLD INTRO we set up a multi-dimensional space S of continuously variable parameterized betting strategies, and then allow a directed search process to explore S for profitable strategies. To evaluate any one strategy, to determine its potential profitability, we have also created a simulator of race-type sporting events and an associated betting exchange on which some number of simulated bettors can interact, thereby creating time-series of synthetic betting-exchange data. Starting from an initial population of randomly-generated betting strategies, we demonstrate here that a simple directed stochastic population-based search, i.e. a form of genetic algorithm (GA), can explore S and identify profitable strategies in usefully short periods of compute-time, tiny fractions of what would be required if an exhaustive brute-force search was used instead.

There is a large array of possible machine learning methods that could be explored in the context of attempting to automatically discover/create profitable strategies for automated gambling. As this project is a proof-of-concept, we decided to first test the base hypothesis that profitable strategies could be discovered via machine learning. To do this, we set up a genetic algorithm (GA) that explores a large continuous space of betting strategies. The space of possible strategies is set up in such a way that we, the designers of the experiment, know in advance that there are points in the space that correspond to known profitable strategies, and the question we then explore is whether a GA can find those known strategies in reasonable time.

In the first instance, as a proof of concept, we have used our betting simulator to create a situation in which a form of betting known as *Dutching* can be profitably deployed. Dutching involves placing multiple bets on different outcomes in the same single event, with the intention of increasing likely returns and/or reducing variance of returns over a sequence of such events. An extension of Dutching, known variously as *Arbitrage Dutch* or *Dutch Arb* can be used in situations where multiple bookmakers are each quoting their own odds for outcomes, and their odds are out of alignment: in such situations a Dutch bet can be placed that exploits the misaligned prices and that is guaranteed to provide a positive return whatever the outcome of the event. The initial research question that our work addresses can thus be phrased in the following terms:

Is it possible to create a simulated system of sports events and bettors gambling on those events which can serve as the platform on which a machine learning (or automated optimization/discovery) process can discover profitable Dutching betting strategies that are transferrable to real-world betting?

IX. Conclusions

This paper has introduced BBE, a simulation of a group of bettors interacting via a betting exchange to make "in play" back and lay bets on the outcome of a race event, while that event is underway. To the best of my knowledge, BBE is the first simulator of its kind, in that no other in-race betting-exchange simulators are available as open-source in the public domain. BBE can be used to generate exhaustively complete data-sets on sub-second temporal resolution from arbitrarily large number of simulated races: this enables the very low-cost generation of extremely large data-sets that can be used for training data-intensive machine learning systems in the search for profitable automated trading systems. Future papers will report on the results from generating and using such BBE-generated data; and, with the BBE source-code made available on GitHub, my hope is that other researchers use BBE as a common platform, facilitating ready replication of results, and also that other researchers contribute to extending the BBE codebase as required.

Acknowledgments

Thanks to Veyndan Stuart, who worked with me on an early proof-of-concept version of the BBE race simulator written in Kotlin when I was advisor on his Masters-thesis project (Stuart, 2019).

References

Abergel, F., Anane, M., Chakrabort, A., Jedidi, A., and Toke, I. (2016) *Limit Order Books*, Cambridge University Press.

Aruajo-Santos, J. P., (2014) *A Trading Agent Framework Using Plain Strategies and Machine Learning.* Master's Thesis, Faculdade de Engenharia da Universidade do Porto.

Axen, G., and Corris, D. (2020) Hedging on Betting Markets. *Risks,* 8(88); doi:10.3390/risks8030088.

BBE (2021) *Bristol Betting Exchange:* open-source code repository on GitHub: www.github.com/ADD-URL-HERE

Benter, W., Miel, G., and Turbough, P. (1996) Modelling Distance Preference in Thoroughbred Racehorses, in de Mestre, N. (editor) *Proceedings of the Third Mathematics and Computers in Sport Conference (3MCS)*, Bond University, pp.207-218.

Betfair (2021) www.betfair.com

Birbeck, E. and Cliff, D. (2018) Using Stock Prices as Ground Truth in Sentiment Analysis to Generate Profitable Trading Systems. *Proceedings of IEEE Symposium on Computational Intelligence in Financial Engineering (CIFEr2018),* November 2018. https://arxiv.org/pdf/1811.02886.pdf

Bouchaud, J., Bonart, J., Donier, J., and Gould, M. (2018), *Trades, Quotes, and Prices*: *Financial Markets Under the Microscope*. Cambridge University Press.

Bradley, I. (2002) The representative bettor, bet size, and prospect theory. *Economics Letters,* 78:409-413.

Brown, A., Rambaccussing, D., Reade, J., and Rossi, G., (2016) *Using Social Media to Identify Market Inefficiencies: Evidence from Twitter and Betfair*. Birkbeck Sport Business Centre Research Paper Series Volume 9(2). Birkbeck University, London.

Brown, A., and Yang, F. (2016) Limited cognition and clustered asset prices: Evidence from betting markets. *Journal of Financial Markets,* 29:27-46.

BSE (2012). *Bristol Stock Exchange:* open-source financial exchange simulator. GitHub repository: https://github.com/davecliff/BristolStockExchange

Bunker, R. and Susnjak, T. (2019) The Application of Machine Learning Techniques for Predicting Results in Team Sport: A Review. Arxiv:1912.1176v1

Camerer, C. (2008) Asset Market Manipulation: A field experiment with racetrack betting. In: *Handbook of Experimental Economics Results*, Volume 1. Elsevier; pp.720-724.

Cameron, C. (2009) *You Bet: The Betfair Story; How Two Men Changed the World of Gambling.* Harper Collins.

Choi, D., and Hui, S. (2014) The role of surprise: Understanding overreaction and underreaction to unanticipated events using in-play soccer betting market. *Journal of Economic Behavior & Organization,* 107:614-629.

Cliff, D. (1997). *Minimal-intelligence agents for bargaining behaviors in market-based environments*. Technical Report HPL-97-91, Hewlett-Packard Labs.

Cliff, D. & Preist, C. (2001) *Days Without End: On the Stability of Experimental Single-Period Continuous Double Auction Markets*, Technical Report HPL-2001-325, Hewlett-Packard Labs.

Cliff, D. (2018) BSE: A Minimal Simulation of a Limit-Order-Book Stock Exchange. In M. Affenzeller, A. Bruzzone, E. Jimenez, F. Longo, Y. Merkureyev, and M. Piera (editors) *Proceedings of the 30th European Modeling and Simulation Symposium* (EMSS2018); pp.194-203.

Cliff, D. (2021) *Technical Note: Parameterized-Response Zero-Intelligence (PRZI) Traders.* Manuscript in preparation.

Das, R., Hanson, J., Tesauro, G., and Khephart, J. (2001). Agent-human interactions in the continuous double auction. Proceedings IJCAI-2001, pages 1169–1176.

Davies, M., Pitt, L., Shapiro, D. and Watson, R. (2005) Betfair.com: Five Technology Forces Revolutionize Worldwide Wagering. *European Management Journal*, 23(5):533-541.

Deffuant, G. (2006). Comparing Extremism Propagation Patterns in Continuous Opinion Models. *Journal of Artificial Societies and Social Simulation,* 9(3):8.

DeGroot, M. (1974). Reaching a Consensus. *Journal of the American Statistical Association*, 69(345):118–121.

De Luca, M. and Cliff, D. (2011a). Agent-human interactions in the continuous double auction, redux. *Proceedings ICAART-2011.*

De Luca, M. and Cliff, D. (2011b). Human-agent auction interactions: Adaptive-aggressive agents dominate. *Proceedings IJCAI-2011.*

De Luca, M., Szostek, C., Cartlidge, J., and Cliff, D. (2011). *Studies of interactions between human traders and algorithmic trading systems.* Driver Review 13, UK Government Office for Science, Foresight Project on the Future of Computer Trading in Financial Markets. http://bit.ly/RoifIu.

Dzalbs, I. and Kalganova, T. (2018) Forecasting Price Movements in Betting Exchanges Using Cartesian Genetic Programming and ANN. *Big Data Research,* 14:112-120.

Easton, S., and Uylangco, K. (2009) Forecasting outcomes in tennis matches using within-match betting markets. *International Journal of Forecasting,* 26:564-575.

Edelman, D. (2008) Modeling Distance Preference and Pace Character in Thoroughbred Turf Racing. Chapter 5 of (Hausch and Ziemba, 2008), pp.67-80.

Edelman, D. (2009) Modelling distance preference in flat racing via average velocity. Chapter 16 of (Vaughan Williams, 2009), pp.339-345.

Elaad, G., Reade, J., and Singleton, C. (2020) Information, prices, and efficiency in an online betting market. *Finance Research Letters,* 35:101291.

Flepp, R., Nuesch, S., and Franck, E. (2017) The liquidity advantage of the quote-driven market: Evidence from the betting industry. *The Quarterly Review of Economics and Finance,* 64:306-317.

Feess, E., Muller, H., and Schumacher, C. (2015) Estimating risk preferences of bettors with different bet sizes. *European Journal of Operational Research,* 249:1102-1112.

Franck, E., Verbeek, E., and Nuesch, S. (2010) Prediction accuracy of different market structures – bookmakers versus a betting exchange. *International Journal of Forecasting,* 26:448-459.

Gjerstad, S. (2003). *The impact of pace in double auction bargaining.* Working paper, Department of Economics, University of Arizona.

Gjerstad, S. and Dickhaut, J. (1997). Price formation in continuous double auctions. *Games and Economic Behavior*, 22(1):1–29.

Gode, D. K. and Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders. *Journal of Political Economy*, 101(1):119–137.

Goncalves, R., Ribeiro, V., Pereira, F., and Rocha., A. (2020) Deep Learning in Exchange Markets. *Information Economics and Policy,* 47:38-51.

Gould, M., Porter., M. Williams, S., McDonald, M., Fenn, D., and Howison, S. (2013) Limit Order Books. *Quantitative Finance,* 13(11):1709-1742. Available from: https://people.maths.ox.ac.uk/porterm/papers/gould-qf-final.pdf

Haahr, F. (2011) *Market Efficiency: An Analysis of the Internet Betting Exchange Market.* MSc Thesis, Copenhagen Business School, Institute of Accounting and Auditing.

Hanke, M., Poulsen, R., and Weissensteiner, A. (2018) Event-Related Exchange-Rate Forecasts Combining Information from Betting Quotes and Option Prices. *Journal of Financial and Quantitative Analysis,* 53(6):2663-2683.

Hausch, D., Lo, V., and Ziemba, W., editors (1994) *The Efficiency of Racetrack Betting Markets.* Academic Press.

Hausch, D., and Ziemba, W., editors (2008) *Handbook of Sports and Lottery Markets.* North Holland.

Hombas, V. and Baloglou, C (1995) Gambling in Greek Antiquity. *Chance,* 18(2):49-50.

Hubáček, O., Sourek, G., and Zelezny, F. (2019) Exploiting sports-betting market using machine learning. *International Journal of Forecasting,* 35:783-796.

Hubáček, O., and Šír, G. (2020) Beating the market with a bad predictive model. https://arxiv.org/pdf/2010.12508.pdf

Hurley, W. and McDonough, L. (1996) The favourite-longshot bias in parimutuel betting: a clarification of the explanation that bettors like to be longshots. *Economics Letters,* 52:275-278.

Hommes, C., and LeBaron, B., (2018) *Computational Economics: Heterogeneous Agent Modeling; Volume 4.* North Holland.

Houghton, J. (2006) *Winning on Betfair for Dummies.* Wiley.

Ioulianou, M., Christofi, I., Talattinis, K., Chalkias, K., and Stephanides, G. (2011) Automatic Identification of Low-Risk Trading Opportunities in Sport Betting Exchange Markets. In Kostoglou, V., Arabatzis, G., and Karamitopolous, L. (editors) *Proceedings of the 1st International Symposium and 10th Balkan Conference on Operational Research*, Volume 1. Pp.108-114.

Iowa Electronic Markets (2021) https://iemweb.biz.uiowa.edu/

Jullien, B., and Salonie, B. (2008) Empirical Evidence on the Preferences of Racetrack Bettors. In Hausch & Ziemba (2008), Chapter 3, pp.27-49.

Kagel, J., and Roth, J., (1997) *The Handbook of Experimental Economics*. Princeton University Press.

Kahneman, D., and Tversky, A., (1979) Prospect Theory: An Analysis of Decision Under Risk. *Econometrica,* 47:273-291.

Kanto, A., Rosenquist, G., and Suvas, A. (1992) On utility function estimation of racetrack bettors. *Journal of Economic Psychology,* 13:491-498.

Liu, B. (2020) *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions.* Second Edition, Cambridge University Press.

Lomas, K. and Cliff, D. (2021) Exploring Narrative Economics: An Agent-Based-Modeling Platform that Integrates Automated Traders with Opinion Dynamics. *Proceedings of the 13th International Conference on Agents and Artificial Intelligence* (ICAART 2021).

Maeda, Y., Tomioka, M., Hanada, M., and Oikawa, M. (2012) Influence of Track Surface Condition on Racing Times of Thoroughbred Racehorses in Flat Races. *Journal of Equine Veterinary Science,* 32:689-695.

Meadows, M. and Cliff, D. (2012). Re-examining the Relative Agreement Model of Opinion Dynamics. *Journal of Artificial Societies and Social Simulation*, 15(4):4.

Meadows, M. and Cliff, D. (2013). The Relative Disagreement Model of Opinion Dynamics: Where Do Extremists Come From? *Proc. 7th International Workshop on Self-Organizing Systems*, (IWSOS), pages 66–77.

Merz, O., Flepp, R., and Franck, E. (2020) *Sonic Thunder vs Brian The Snail: Are people affected by uninformative racehorse names?* Working Paper No. 384, University of Zurich Department of Business Administration.

Mitra, G. and Yu, X. (2016) *Handbook of Sentiment Analysis in Finance.* Optirisk.

O'Leary, D. (2012) Internal Corporate Prediction Markets: From each according to his bet. *International Journal of Accounting Information Systems,* 14:89-103.

Palomino, F., Renneboog, L., and Zhang, C. (2009) Information salience, investor sentiment, and stock returns: the case of British soccer betting. *Journal of Corporate Finance,* 15:368-387.

Pentapalli, M. (2008). *A comparative study of Roth-Erev and Modified Roth-Erev reinforcement learning algorithms for uniform-price double auctions.* PhD thesis, Iowa State University.

Piron, R., and Smith, L. (1995) Testing risklove in an experimental racetrack. *Journal of Economic Behavior and Organization*, 27:465-474.

Plott, C., & Smith, V., editors (2008) *Handbook of Experimental Economics Results, Volume 1.* North-Holland.

Pozzi, F., Fersini, E., Messina, E., and Liu, B. (2016) *Sentiment Analysis in Social Networks.* Morgan Kauffman.

PredictIt (2021) www.predictit.org

PRNewswire (2021) https://www.prnewswire.com/news-releases/global-gambling-industry-301107436.html

Rodríguez, P., Humphreys, B., and Simmons, R., editors (2017) *The Economics of Sports Betting.* Edward Elgar.

Restocchi, V. (2018) *It Takes All Sorts: The Complexity of Prediction Markets*. PhD Thesis, University of Southampton.

Roy, P. (2010) *British Horseracing Reacts to Betfair's IPO Announcement.* Press release from Chairman of the British Horseracing Authority, 21st September 2010. https://www.britishhorseracing.com/press_releases/british-horseracing-reacts-to-betfairs-ipo-announcement/

Rust, J., Palmer, R., and Miller, J. (1992). Behaviour of trading automata in a computerized double auction market. In: *The Double Auction Market: Theories and Evidence*, pages 155–198. Addison Wesley.

Shiller, R. (2017) *Narrative Economics.* Technical Report 2069, Cowles Foundation, Yale University.

Shiller, R. (2019) *Narrative Economics: How Stories Go Viral and Drive Major Economic Events.* Princeton University Press.

Smarkets (2021) www.smarkets.com

Smith, M., and Vaughan Williams, L. (2008). Betting Exchanges: A Technological Revolution in Sports Betting. Chapter 19 of (Hausch and Ziemba, 2008), pp403-418.

Smith, M., and Vaughan Williams, L. (2010) Forecasting horse race outcomes: New evidence on odds bias in UK betting markets. *International Journal of Forecasting,* 26:543-550.

Smith, V., 1962. An Experimental Study of Competitive Market Behavior. *Journal of Political Economy* 70(2):111-137.

Smith, V., editor (2000) *Bargaining and Market Behavior: Essays in Experimental Economics.* Cambridge University Press.

Stotter, S., Cartlidge, J., and Cliff, D. (2014) Behavioural investigations of financial trading agents using Exchange Portal (ExPo). In: N. T. Nguyen, R. Kowalczyk, A. Fred, & F. Joaqui, editors: *Transactions on Computational Collective Intelligence XVII*. Berlin Heidelberg: Springer, pp. 22-45.

Stuart, V. (2019) *Towards the Simulation of a Betting Exchange: Designing and implementing a betting exchange with automated bettors betting on simulated horse races*. MEng Thesis, Department of Computer Science, University of Bristol.

Suhonen,N., Saastamoinen, J., Kainulainen, T., and Forrest, D. (2018) Is timing everything in horse betting? Bet amount, timing and bettors' returns in pari-mutuel wagering markets. *Economics Letters,* 173:97-99.

Sung, M., and Johnson,. J. (2008) Semi-Strong Form Information Efficiency in Horse Race Betting Markets. Chapter 14 of Hausch & Ziemba (2008), pp.275-306.

Sung, M., Johnson, J., and McDonald, D. (2016) Informed trading, market efficiency, and volatility. *Economics Letters,* 149:65-59.

Swidler, S., and Shaw, R. (1995) Racetrack Wagering and the "Uninformed" Bettor: A Study of Market Efficiency. *The Quarterly Review of Economics and Finance,* 35(3):305-314.

Tesauro, G. and Bredin, J. L. (2002). Strategic sequential bidding in auctions using dynamic programming. In *Proc. First Int. Joint Conf. on Autonomous Agents and Multiagent Systems: Part 2*, pp.591–598.

Tesauro, G. and Das, R. (2001). High-performance bidding agents for the continuous double auction. *Proc. 3rd ACM Conference on E-Commerce*, pp.206–209.

Tversky, A., and Kahneman, D., (1992) Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty,* 5: 297-323.

Thorp, E. (2006) The Kelly Criterion in Blackjack Sports Betting, and the Stock Market. In Zenios, S., and Ziemba, W. (editors) *Handbook of Asset and Liability Management*, Volume 1. Elsevier; pp.385-428.

Thompkins, R., Ziemba, W., and Hodges, S. (2008) The Favourite-Longshot Bias in S&P500 and FTSE100 Index Futures Options: The Returns to Bets and the Cost of Insurance. In Hausch & Ziemba (2008), pp.161-180.

Tsrimpas, p. (2015) *Specification and Performance Optimization of Real-Time Trading Strategies for Betting Exchange Platforms.* PhD thesis, Imperial College London, Department of Computing.

Vaughan Williams, L. (editor) (2003) *The Economics of Gambling*, Routledge.

Vaughan Williams, L. (editor) (2009) *Information Efficiency in Financial and Betting Markets*, Cambridge University Press.

Vytelingum, K., Cliff, D., and Jennings, N. (2008). Strategic bidding in continuous double auctions. Artificial Intelligence, 172(14):1700–1729.

Wall, M., Costello, R., and Lindsay, S. (2017) The miracle of thje markets: Identifying key campaign events in the Scottish Independence Referendum using betting odds. *Electoral Studies,* 46:39-47.

Wheatcroft, E. (2020) A profitable model for predicting the over/under market in football. *International Journal of Forecasting,* 36:916-932.

Wilkens, S. (2020) Sports Prediction and Betting Models in the Machine Learning Age: The Case of Tennis. *SSRN3506302.*