On Fairness in Continuous Electronic Markets

Hayden Melton*
Thomson Reuters (Markets) LLC
3 Times Square
New York, NY 10036
hayden.melton@tr.com

ABSTRACT

Most of the world's financial markets are *electronic* (i.e., are implemented as software systems) and *continuous* (i.e., process orders received from market participants immediately, on a FIFO basis). In this short position paper I argue that such markets cannot provide 'racetrack fairness' to their participants, yet this form of fairness seems to feature quite prominently throughout the large, multi-jurisdictional body of law governing financial markets. What seems to follow from this is that electronic *batch-style* markets are not only a *desirable* replacement for continuous ones—as a number of economists have recently argued—but a *necessary* replacement.

1 INTRODUCTION

Electronic trading venues (ETVs) are widely-deployed software systems that, on a daily basis, facilitate the exchange of trillions of dollars in the world's financial markets [14]. These ETVs can be characterized on the basis of whether they implement a *batch-* or *continuous*-style of market [10]. In the former order messages received by the ETV are subject to the deliberate imposition of delays after their receipt by the venue, so a batch of such orders can first accumulate, and later be jointly processed by the venue; in the latter orders are processed by the venue immediately upon their receipt—so serially, in a first-in-first-out (FIFO) manner.

Economists, in noting that most of the world's financial markets are both electronic and continuous, have recently identified what many perceive as a problem of fairness in these markets.[1] In particular, they have described a technology 'arms race' among market participants where massive and ongoing investments are being made solely in pursuit of 'speed', and worse, to attain seemingly only miniscule increases in it [4, 5, 9, 20]. It is, of course, the FIFO processing of continuous markets that are causing participants to invest in speed-enhancing technology, because when a profitable trading opportunity manifests on an ETV, among the participants competing to capture that opportunity it will be the fastest participant that succeeds in doing so, because that participant's order will reach (and be processed by) the ETV first. These seemingly miniscule increases in speed—and recent evidence suggests that nanoseconds may separate the 'fast' from the 'slow' [13]—are of course an artifact of the (ever increasing) speed at which computers and computer networks operate.[2]

What seems to be an implicit assumption in the works of economists on this 'arms race' is that the ETVs implementing these continuous markets are themselves 'fair'. Much like the conclusion 'the fastest runner in a race will win it' implicitly assumes that the race itself is fair, i.e., none of the runners in it are receiving a head-start or getting to run a shorter track, the assumption is that the software system that *is* the ETV is capable of ensuring participants receive market data updates sent by it at the same time, and that order messages are ultimately processed in the same (temporal) ordering in which they were received by it.[3] Unfortunately, and as is described in detail elsewhere [6, 12–14], it is a non-trivial engineering undertaking to ensure ETVs are fair in this regard. Indeed, over the past several years controversies pertaining to this form of fairness on major ETVs worldwide have been the subject of significant coverage in even mainstream media [14].

The question this paper seeks to address is whether or not an ETV implementing a continuous market can ever truly be fair, where the form of fairness under consideration is that analogized earlier with a running race. If the answer to this question is 'no' then besides strengthening the recent position of economists that continuous electronic markets are 'bad' [5, 9], it would also seem to have important policy implications since there is a large, multi-jurisdictional body of law seemingly requiring this form of fairness in financial markets [14]. The novelty of the arguments made in this paper for batch-style markets in pursuit of fairness is that

*Views expressed herein do not constitute legal or investment advice and do not necessarily reflect those of the author's employer.

[1] References to those who perceive this as a problem of fairness are cited elsewhere [14], but ultimately their collective perception of it boils down to what has been termed '(in)equality of outcomes' [1]. The outcomes seem unjust because only the largest market participants can afford to make the massive investments in speed-enhancing technology required to capture profitable trading opportunities, leaving smaller participants disenfranchised.

[2] In this context *speed* is *units work performed per unit time*. If work is held constant, and a technological advancement (e.g., as predicted by Moore's Law) causes speed to double, then the time taken to perform the same unit of work will halve. Under this regime—if speed is measured in units time—it doesn't take long for the meaningful scale of measure to go from milliseconds to microseconds, to nanoseconds. In light of this math one must wonder about the wisdom in the concern that these increases in speed are only miniscule.

[3] A lengthier exposition of this analogy appears elsewhere [12], but to summarize: (i) variance in the sending times of market data updates by the venue to participants correspond to headstarts, (ii) variance in order processing times by the venue which cause violations in the FIFO discipline correspond to varying lengths of racetrack, and (iii) the time taken for a participant to respond to their market data update with an order corresponds to that participant's speed (and maximizing this speed relative to other participants is the 'arms race' described by economists). It is the *sum* of the times taken for (i), (ii) and (iii) then that determine which participant wins a race for a trading opportunity on an ETV, and it is this sum that makes clear the relationship between the works of economists on this 'arms race' and the form of fairness addressed in this paper.

FairWare'18, May 29, 2018, Gothenburg, Sweden

ACM ISBN 978-1-4503-5746-3/18/05…$15.00
https://doi.org/10.1145/3194770.3194771

they apply even in this age of co-location—where all participants' trading systems can be placed in the same datacenter as the ETV. Historical arguments made to the same end are weaker because they have required participants to be geographically distributed and thus subject to differing propagation delays in their transmission of data to and from the ETV [8, 23]. Further, if one subscribes only to the dictum of *equal opportunity* of access in financial markets [1], in light of co-location these historical arguments may not pertain to fairness at all.

2 WHY CONTINUOUS ELECTRONIC MARKETS AREN'T 'RACETRACK FAIR'

Responsive to massive and ongoing increases in order traffic sent by electronic participants to ETVs—and so as to ensure consistent and small response times in processing those orders—software engineers have historically architected ETVs to parallelize the operations they perform [14]. An exposition of this style of ETV architecture and its consequences is provided elsewhere [14, §II-B], but what is important to note here is that this architecture is a major cause of the *racetrack unfairness* in our analogy. As this problem has become more widely-known many ETVs have sought to address it through various software changes (as later shown in Fig.1) aimed at guaranteeing FIFO processing of orders [6][7, p.3].

If one assumes that an ETV implementing a continuous market can ensure that each market data update is distributed to all interested market participants at effectively the same time (e.g., because all such participants are co-located in the same datacenter as the ETV, and all cables connecting each participant to the ETV are the same length, media and bandwidth, and because IP multicast is used in combination with a switch implementing the appropriate discipline for same-time output across ports) and that it actually processes order messages in a FIFO manner, then can we consider that ETV (racetrack) fair? There is an argument that we still *cannot*, but a prerequisite to the exposition of that argument is a discussion of 'tie-breaking' when two or more orders are received by the ETV at effectively the same time.

Since the resources corresponding to the profitable trading opportunities for which participants on an ETV compete—usually queue position in the limit order book when *price-making*, and the competitively priced bids and offers it contains when *price-taking* [13]—are *not* infinitely divisible[4] a discussion of what constitutes a 'fair division' of each such opportunity is required. Put another way, if multiple orders all competing for a single (indivisible) resource are received by the ETV simultaneously, how should we perform tie-breaking to assign it to one of those orders? Economists have quite extensively studied problems of this general form (see e.g., [3, 16]) and they have concluded that achieving fairness therein requires the use of a 'lottery' to allocate the resource, and crucially one that ensures the 'equal treatment of equals'. In our specific context of racetrack fairness the entities competing are market participants (and *not* orders)[5], and two or more such participants

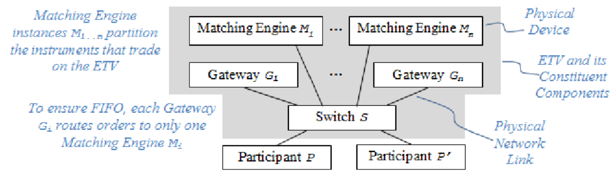

Figure 1: An architecture for guaranteeing FIFO.

are equally matched if they have all made the same investments in speed-enhancing technology (so are all equally as fast as one-another). Thus fairness in tie-breaking here means each equally-fast participant competing for the trading opportunity should receive exactly one lottery ticket for it and thus have an equal probability of being allocated it.

In light of the preceding discussion of what racetrack fairness means in the presence of ties and indivisible resources, consider now the ETV of Fig.1 that has been architected to guarantee FIFO order processing (as in the styles of [6, 7]). In the figure, the computers of market participants P and P' send orders to the computers of the ETV hosting its order gateways components $G_{1..n}$ via the ETV's network switch S. Each such gateway G_i then forwards those orders to the computer hosting the ETV's corresponding matching engine component M_i again via S. Each M_i performs FIFO allocation of trading opportunities to orders, and the totality of the financial instruments that trade on the ETV are partitioned across $M_{1..n}$. Except for in S—where its output to a single physical link can cause a *total ordering* of packets that did not exist at its input from distinct market participants via multiple physical links[6]—all other components of the ETV are implemented so as to each individually guarantee FIFO in their own processing.

What should be clear of the ETV in Fig.1 now is that it is the network switch S that has exclusive responsibility for tie-breaking when two or more orders are received by the ETV at the same time. The question of fairness thus becomes one of whether the switch, in performing that tie-breaking, is capable of assigning one lottery ticket per participant per trading opportunity. In the general case the answer to that question is very likely 'no' because switches are general purpose networking devices that do not have knowledge of *application-level* data (i.e., each specific trading opportunity, each participant, and their individual orders)—rather they are designed to operate only at lower-levels in the OSI model of the network stack (i.e. to only have knowledge of things like ports and packets).

Perhaps worse than the switch acting as arbitrator of ties though is the serialization of messages that must occur on each participant's physical link to the ETV. When this serialization is considered in combination with the observation that a single event can sometimes trigger races among participants for a plurality of different trading opportunities, the likelihood of the ETV in Fig.1 providing fairness diminishes even further. To understand why, consider a participant seeking to capture the entire plurality of opportunities indicated by a single event. To do this the participant will need to send at least one order for each opportunity, and consequently will end

[4] Even to the extent that a single trading opportunity may be divisible (e.g., because it comprises multiple units of the instrument), treating that trading opportunity as indivisible may result in better outcomes for all market participants [17].

[5] The distinction is important because if 'lottery tickets' were instead issued on a per order (and *not* per participant) basis, then to improve their chances of being allocated the trading opportunity—but to the detriment of the ETV's fairness—a participant could send multiple redundant copies of the same order to it. This exact behavior has been observed on a real-life ETV [6].

[6] This is more generally known as *output port contention*. When two or more input ports receive packets simultaneously, and when those packets are all destined for the same output port, the switch must decide a total ordering in which to send those packets because only one can be sent on the (single) output link at a time [19, p.103].

up sending either a much larger (network) packet or many more packets on his link than any other participant that is competing for just a single one of the opportunities will send on her link. The end result is that more active participants are disadvantaged because serialization will cause some of their orders arrive later at the switch (and thus later at ETV) than those of the less active participants with which they are competing. This is contrary to fairness because it will happen even if all the participants have made the same investments in speed-related technology (so are all equally fast).[7]

3 CONCLUSION

When specific behaviors of market participants described in the previous section are viewed through the lens of computer networking technology, the answer to the question of whether ETVs implementing continuous markets can be considered *racetrack fair* appears to be 'no'. The fact that all these specific behaviors have been observed on a major, real-world ETV [13, 14] gives weight to the view that the arguments provided here have a sound practical basis and are not mere intellectual curiosities. The policy implication that seems to follow is that batch-style markets are not merely a *desirable* replacement for continuous electronic markets (as in [5, 9]), they are a *necessary* one.

An interesting question that naturally arises from this work then is: why, if continuous electronic markets are unfair, do they remain predominant form of financial market?[8] One answer may be that the arguments provided in this paper are not widely-known. Another may be that changes to market microstructure such as tick size reduction on a venue may cause (market-making) participants to compete more on 'price' than 'speed', thereby reducing the extent of the problem on that venue [24]. Another answer may be that many venue operators are unwilling to slow their markets from continuous to batch-style because they derive a large proportion of their revenue from selling 'speed' to market participants [4]. Yet another answer may be that the regulatory requirements in certain markets, such as in US equities, while requiring fairness also generally prohibit the deliberate imposition of delays on orders by ETVs [22]. Lastly—and much like there often exists trade-offs to be made between fairness and efficiency [2]—one must also consider the possibility that the overall benefits of continuous electronic markets outweigh the 'racetrack unfairness' they cause (e.g., because the increases in speed they have enabled have improved trading risk management [21, p.28]).

[7] What may seem an easy solution to this problem—allocating a participant one physical link for each type of trading opportunity he may compete for—is impractical for at least two reasons. First, there are an unbounded number of distinct types of trading opportunities, yet there are limits on the number of ports (and therefore physical links) a commercial switch can support. Second, and to the detriment of fairness, assignment of multiple physical links to a single participant may incentivize her to send redundant copies of the same order message in a race for a single trading opportunity, in an attempt to increase her probability of winning it (as has been observed in [6]).

[8] As part of the exercise described in [13]—which involved the conversion of a major, long-lived ETV from a continuous to batch-style market—feedback on this change was sought from participants on the venue. A sentiment expressed by several such participants, as if it were axiomatic, was that "the *only* fair way to process orders is FIFO; batching is unfair". A reason for this strong negative view of batching may be that in queuing systems in general, non-FIFO processing has been perceived as socially unjust [11]. This perception may date as far back as 1670AD, as evidenced by the old English proverb *the early bird gets the worm* [18].

One final point warrants mention. It has been shown elsewhere that the extent to which a batch-style market achieves or approaches 'racetrack fairness' is highly dependent on the design of the buffering mechanism it uses to impose the delays on orders, and the length of the delays imposed [12–14]. What this means is that the mere classification of an ETV as being a batch-style market is *not* synonymous with it providing 'racetrack fairness'. To the extent that the delay imposed exceeds that required to achieve 'racetrack fairness' on an ETV, a minimum response time for participants is established [15, ¶65], and the 'arms race' for speed on that venue is abated.

ACKNOWLEDGEMENTS

The author thanks the anonymous reviewers for their detailed comments and useful suggestions that have been used to improve the paper.

REFERENCES

[1] James J Angel and Douglas McCabe. 2013. Fairness in financial markets: The case of high frequency trading. *Journal of Business Ethics* 112, 4 (2013), 585–595.
[2] Dimitris Bertsimas, Vivek F Farias, and Nikolaos Trichakis. 2012. On the efficiency-fairness trade-off. *Management Science* 58, 12 (2012), 2234–2250.
[3] Anna Bogomolnaia and Hervé Moulin. 2001. A new solution to the random assignment problem. *Journal of Economic theory* 100, 2 (2001), 295–328.
[4] Eric Budish. 2017. AEA/AFA Joint Luncheon - Will the Market Fix the Market? https://goo.gl/dWEiCa. (Jan. 6 2017). Accessed Jan 8 2017.
[5] Eric Budish, Peter Cramton, and John Shim. 2015. The high-frequency trading arms race: Frequent batch auctions as a market design response. *Q. J. Econ* 130, 4 (2015), 1547–1621.
[6] CME Group. 2014. Slides from New iLink Architecture Webinar (Part I). https://goo.gl/mWdDa9. (2014). Accessed Sep 4 2016.
[7] Eurex. 2017. Insights into trading system dynamics: Eurex Exchange's T7. https://goo.gl/wySDaP. (March 2017). Accessed Dec 31 2017.
[8] Dhananjay K Gode and Shyam Sunder. 2000. Designing electronic markets: on the impossibility of equitable continuously-clearing mechanisms with geographically distributed agents. https://goo.gl/dx2wpi. (2000). Accessed Jan 15 2018.
[9] Larry Harris. 2013. What to do about high-frequency trading. *Financial Analysts Journal* 69, 2 (2013).
[10] Roger D Huang and Hans R Stoll. 1992. The design of trading systems: Lessons from abroad. *Financial Analysts Journal* (1992), 49–54.
[11] Richard C Larson. 1987. OR forum–perspectives on queues: Social justice and the psychology of queueing. *Operations research* 35, 6 (1987), 895–905.
[12] Hayden Melton. 2017. A fairness-oriented performance metric for use on electronic trading venues. In *CSCI'17*. IEEE, Las Vegas NV, 1027–1030.
[13] Hayden Melton. 2017. Market Mechanism Refinement on a Continuous Limit Order Book Venue: A Case Study. *SIGecom Exch.* 16, 1 (Sept. 2017), 72–77.
[14] Hayden Melton. 2017. Understanding and improving temporal fairness on an electronic trading venue. In *ICDCSW*. IEEE, Atlanta GA, 1–6.
[15] Hayden P Melton. 2015. Systems and methods for quantifying temporal fairness on electronic trading venues. US Patent App. 14/930,499. (Nov. 2 2015).
[16] Hervé Moulin. 2004. *Fair division and collective welfare.* MIT press.
[17] Roel Oomen. 2017. Execution in an aggregator. *Quant. Fin.* 17, 3 (2017), 383–404.
[18] John Ray. 1670. *A collection of English proverbs.* John Hayes, Cambridge, UK. Available via https://books.google.com/books?id=NhUsAQAAMAAJ.
[19] Roberto Rojas-Cessa. 2016. *Interconnections for Computer Communications and Packet Networks.* CRC Press.
[20] Robert A Schwartz. 2010. Comment Letter to the SEC: Regulation of Dark Pools and Other Trading Interests (File No. S7-27-09). https://goo.gl/6sis8t. (Feb. 22 2010). Accessed Mar 11 2018.
[21] Robert A Schwartz, John Aidan Byrne, and Eileen Stempel. 2017. *Rapidly Changing Securities Markets: Who Are the Initiators?* Springer.
[22] Securities and Exchange Commission. 2016. In the Matter of the Application of Investors' Exchange, LLC for Registration as a National Securities Exchange: Findings, Opinion, and Order of the Commission (Release No. 34-78101; File No. 10-222). https://goo.gl/cjzJtE. (2016). Accessed Mar 6 2018.
[23] Charles Tresser and Daniel Sturman. 2001. Fair and scalable trading system and method. (May 23 2001). US Patent App. 09/864,015.
[24] Mao Ye. 2017. Who provides liquidity and when: An analysis of price vs. speed competition on liquidity and welfare. (2017). Available at *SSRN 2900521*.