

October 21, 2019

Vanessa Countryman, Secretary  
Securities and Exchange Commission  
100 F Street NE  
Washington DC

Re: Release 34-87096; File Number SR-CboeEDGA-2019-012; Cboe EDGA Exchange, Inc. ("EDGA"); Order Instituting Proceedings to Determine Whether to Approve or Disapprove a Proposed Rule Change to Introduce a Liquidity Provider Protection Delay Mechanism on EDGA ("Order")

Dear Ms. Countryman:

This is in response to the questions included in the Order.<sup>1</sup> I reproduce the questions below in italics and follow each one with commentary.

*1. Do commenters agree with the Exchange's assertion that the proposal would reduce cross-market latency arbitrage and improve market quality by enabling liquidity providers to maintain tighter spreads for longer durations and with greater size? Why or why not? How should enhancements to market quality be measured?*

EDGA hasn't shown that its market suffers from cross-market latency arbitrage. EDGA hasn't offered convincing evidence to support the story in its rule filing that arbitrageurs observe price discovery in the Chicago futures markets, leverage expensive microwave technology to outrace other firms to its matching engines in New Jersey, and then pick off stale orders on its book.<sup>2</sup> The little evidence it has submitted<sup>3</sup> is confusing and contradictory. Instead of supporting EDGA's narrative that evidence might show that any cross-market latency arbitrage is benign<sup>4</sup> and could suggest an altogether different effect of the Filing.<sup>5</sup>

In a real-world experiment of a discriminatory speed bump very similar to EDGA's proposal, Canadian regulators found that TSX Alpha didn't produce any material improvements in market quality measures like spreads, price impact, and quoted depth, but did show higher costs for institutional investors and increased market complexity.<sup>6</sup> The only academic study on that experiment found that the discriminatory speed bump produced a discriminatory effect because it "segments order flow and increases profits for fast liquidity providers on that venue at the expense of other liquidity providers and aggregate market quality."<sup>7</sup> Those increased profits are a regulatory subsidy without any corresponding benefit to other market participants. If adopted market-wide, on conservative assumptions, I estimated the value of that subsidy in the U.S. ranges to \$1 billion or more per year, while on EDGA alone the subsidy could be \$75,000 or more per day.<sup>8</sup> Investors will pay for that subsidy when they attempt but fail to trade with advertised prices and pay worse prices instead.

---

<sup>1</sup> Order, pages 41-43.

<sup>2</sup> Release 34-86168; File Number SR-CboeEDGA-2019-012; Cboe EDGA Exchange, Inc.; Notice of Filing of a Proposed Rule Change to Introduce a Liquidity Provider Protection on EDGA ("Filing"), pages 4-6.

<sup>3</sup> Letter from Adrian Griffiths, Assistant General Counsel, Cboe, to Vanessa Countryman, Secretary, Securities and Exchange Commission, August 22, 2019 ("Cboe Letter").

<sup>4</sup> Letter from R. T. Leuchtkafer to Vanessa Countryman, Secretary, Securities and Exchange Commission, September 9, 2019 ("Leuchtkafer 1"), pages 2-4.

<sup>5</sup> Leuchtkafer 1, page 6.

<sup>6</sup> See the research discussion in Leuchtkafer 1, page 7.

<sup>7</sup> Chen, Foley, Goldstein, and Ruf, "The Value of a Millisecond: Harnessing Information in Fast, Fragmented Markets," (2017), available at <https://www.ssrn.com/abstract=2860359> ("Chen"), page 1.

<sup>8</sup> Leuchtkafer 1, page 5.

2. According to several commenters, EDGA liquidity would be “illusory” because the Exchange’s liquidity providers could update their quotations while incoming orders are delayed. Do commenters believe that the proposed rule change would lead to quote fading? Why or why not? Do commenters believe that the proposed rule change would impact fill rates? Would the “illusory” liquidity be a significant portion of the Exchange’s overall liquidity?

The purpose of discriminatory speed bumps is quote fading. That's their explicit intent. Canadian regulators said so very plainly about their own marketplace's discriminatory speed bump, "quote fade should be expected on Alpha, given that its [discriminatory speed bump] model enables providers of liquidity to cancel their orders in response to order book changes on Alpha and/or other marketplaces."<sup>9</sup>

Markets today are infested with illusory liquidity. The idea of a *bona fide* order is a quaint anachronism. Various studies have shown that even in unbiased distributed marketplaces - fragmented marketplaces without discriminatory speed bumps, in other words - more than 40% of HFT displayed orders fade after a trade.<sup>10</sup> In discriminatory marketplaces the problem is much worse. The academic study of the speed bumped TSX Alpha in Canada estimated quote fading on Alpha at 66% on average and in some circumstances at 88% or higher.<sup>11</sup>

.....

3. Some commenters assert that the proposal is not unfairly discriminatory under the Exchange Act because the proposal addresses a particular behavior as opposed to specific class or type of market participants. Is this assertion accurate? Why or why not?

The Filing and Cboe Letter don't demonstrate that the behavior - cross-asset arbitrage - exists on EDGA's marketplace, or that if it does exist that it produces any material harm to its market.<sup>12</sup>

But suppose for a moment that the Filing and Cboe Letter have demonstrated a material harm. Following EDGA's cross-asset arbitrage narrative, here's a list of resources a market participant likely needs to have in place to take advantage of EDGA's proposed remedy:

- Chicago Mercantile Exchange high-speed data feeds;
- Co-located servers at CME;
- High-speed networks between Chicago and New Jersey;
- EDGA high-speed data feeds;
- Co-located servers in EDGA's data center;
- Data feed handlers, order handlers, software stacks to build and maintain exchange order books, order and trade management software, algorithms to calculate fair value and generate orders, risk management software, clearing firm interfaces;
- The necessary development staff to design, build, and maintain all proprietary software;
- The necessary procurement and vendor management capabilities to implement and maintain exchange, network provider, and other contracts;
- Appropriate registrations and registered staff, compliance staff, and counsel;
- The necessary operations capabilities to install, access, and maintain trading infrastructure.

While it's true that EDGA in its majestic equality will empower both retail investors like Mrs. Betty Johanssen of Red Lake, Minnesota and firms like XTX Markets to take advantage of its proposed speed

---

<sup>9</sup> Ontario Securities Commission Staff Notice 21-712 ("OSC Notice") available at [https://www.osc.gov.on.ca/documents/en/Securities-Category2/20180202\\_21-712\\_sn-alpha-impact.pdf](https://www.osc.gov.on.ca/documents/en/Securities-Category2/20180202_21-712_sn-alpha-impact.pdf), page 3.

<sup>10</sup> European Securities and Markets Authority, "Order duplication and liquidity measurement in EU equity markets" (2016), page 18.

<sup>11</sup> Chen, pages 15-16 (see Table 1).

<sup>12</sup> See note 4.

bump, what class of market participants do we imagine is the certain *de facto* beneficiary? The SEC can't be blind to this and let EDGA shelter behind a claim of equal opportunity. The Filing has a veneer of nondiscrimination while landing far short of universal utility. Its representations that it will benefit a "wide range of orders" can only make us all wonder at just how narrowly EDGA understands "wide."

If it looks like a market maker subsidy, walks like a market maker subsidy, and quacks like a market maker subsidy, it's a market maker subsidy.

.....

*4. Will the proposal increase the risk of adverse selection for liquidity takers and market participants that are unable to react to market signals in order to adjust their quotes within four milliseconds?*

By enabling a privileged class to flee adverse selection, the proposal shifts the costs of adverse selection onto participants who leave resting orders on Cboe EDGA but can't leverage the speed bump. It will also shift those costs to exchanges which don't have a discriminatory speed bump.

Speed bumps should at least pass a simple Three Stooges test: If Moe throws a punch at Curly, Curly ducks, and Moe hits Larry on the follow-through, Curly is spared but Larry gets socked.<sup>13</sup> Since discriminatory speed bumps like this only change who gets socked, they fail the test.

The most affluent and sophisticated market makers can fade while institutions, retail customers, and less affluent professionals stay exposed, and are then disproportionately adversely selected. EDGA's proposal doesn't cure latency arbitrage, real or imagined, it merely shuffles who suffers from it, and ironically the victims here will be market participants who don't use the fastest and most expensive technology and exchanges who don't deploy a discriminatory speed bump. And as a principle, discriminatory delays explicitly or implicitly favoring resource rich market intermediaries will increase the value of all their other regulatory subsidies while they shift costs to, extracting rents from, unsubsidized and resource poor market participants. I've estimated those rents at \$75,000 a day on EDGA and \$1 billion a year if adopted by the market as a whole.<sup>14</sup>

.....

*5. Is an intentional delay of four milliseconds necessary to minimize the effectiveness of latency arbitrage strategies? Will the delay negate the advantages that trading firms using the latest microwave connections have over liquidity providers using traditional fiber connections? Should the delay be shorter or longer to accomplish this goal? Is four milliseconds an appropriate duration for a delay? Is such delay consistent with the Act? Why or why not?*

EDGA hasn't shown that its market suffers from cross-market latency arbitrage. EDGA hasn't offered convincing evidence to support the story in its rule filing that arbitrageurs observe price discovery in the Chicago futures markets, leverage expensive microwave technology to outrace other firms to its matching engines in New Jersey, and then pick off stale orders on its book.<sup>15</sup> The little evidence it has submitted<sup>16</sup> is confusing and contradictory. Instead of supporting EDGA's narrative, that evidence might show that any cross-market latency arbitrage is benign<sup>17</sup> and could suggest an altogether different effect of the Filing.<sup>18</sup>

---

<sup>13</sup> See a clip of Moe, Larry, and Curly at <https://www.dailymotion.com/video/x4yvnf2?start=150>.

<sup>14</sup> Leuchtkafer 1, page 5.

<sup>15</sup> Filing, pages 4-6.

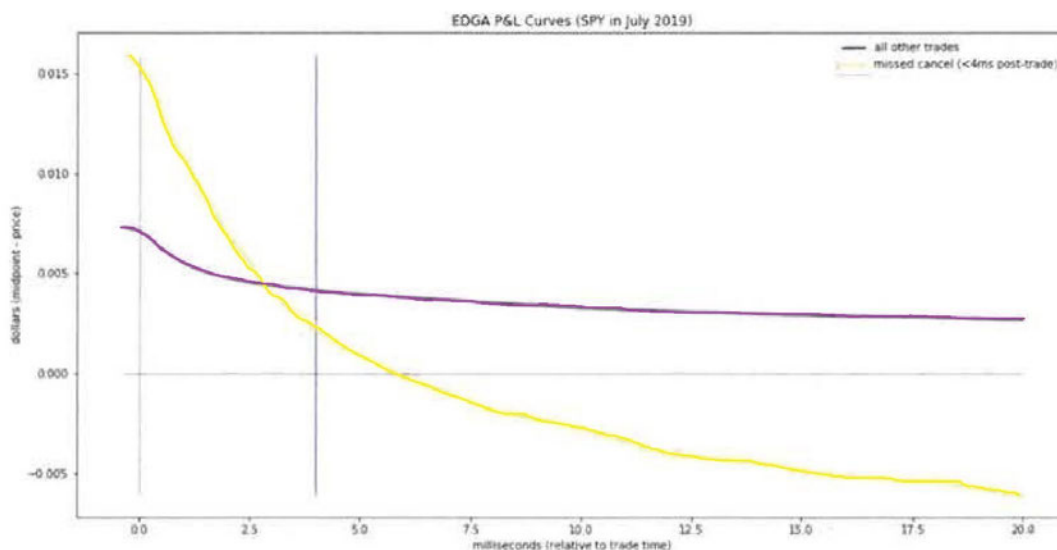
<sup>16</sup> Cboe Letter, pages 3-4.

<sup>17</sup> Leuchtkafer 1, pages 2-4.

<sup>18</sup> Leuchtkafer 1, page 6.

## Paradox

Although discussed at length in Leuchtkafer 1, let's also go through EDGA's core evidence here. EDGA describes the following graph as showing "markouts for liquidity providers on EDGA in SPY during the month of July 2019 based on whether or not the transaction involved a 'missed cancel' - i.e. where the liquidity provider attempted and failed to cancel or replace their quotation within four milliseconds after an execution." The yellow and purple lines ("Yellow Sample" and "Purple Sample" respectively) show "the difference between the execution price and the midpoint price at the time of the trade and in the milliseconds following an execution." According to EDGA, the Yellow Sample shows markouts where a market maker attempted but failed to cancel an order before execution. EDGA points out that prices move dramatically after the trades. The Purple Sample shows markouts for all other trades - presumably where the market maker didn't attempt to cancel the order before execution. EDGA says prices here are relatively stable.<sup>19</sup>



Since the Yellow Sample<sup>20</sup> selection criteria condition on a failed market maker cancel, if EDGA's narrative in the Filing is correct, we can assume at  $t_0$  in the graph (trade time) we are four milliseconds along from the relevant event in Chicago. According to the Filing, that's because there is a group of superfast latency arbitrageurs using high-speed microwave connections to get to EDGA's data centers in New Jersey from the futures markets in Chicago in as little as four milliseconds ("Chicago Arbs").

On the other hand, as EDGA described it in the Filing, its market makers get their signals from Chicago in about eight milliseconds because they use slow fiber. Since its market makers rely on those slower signals from Chicago before trying to cancel, it will take at least another four milliseconds after  $t_0$  before market makers will even know they should cancel their quotes.<sup>21</sup> And then their cancels fail because the Chicago Arbs have already traded with their quotes.

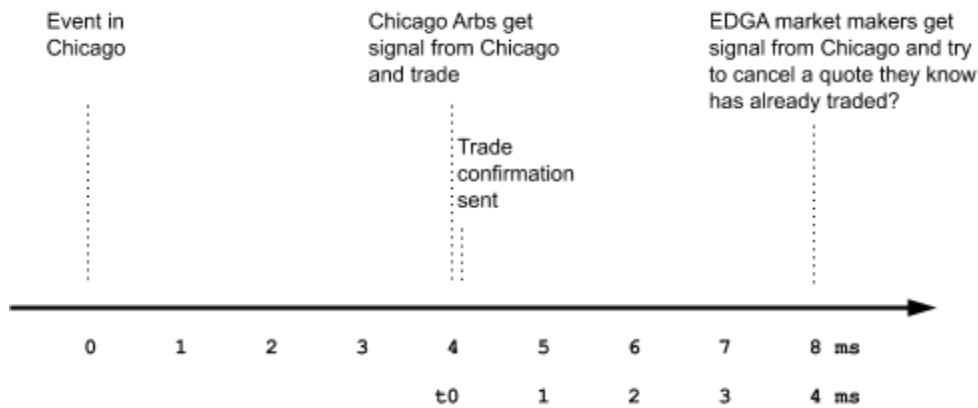
Note the apparent paradox. Why is the market maker trying to cancel an order it surely *already knows* was traded out almost four milliseconds ago - a relative eternity - by the Chicago Arbs, by the firms that

<sup>19</sup> Cboe Letter, pages 3-4.

<sup>20</sup> All images have been enhanced to make the yellow and purple lines more legible.

<sup>21</sup> See the discussion of microwave and fiber and their associated latencies in the Filing at page 6.

got to EDGA so much faster because they use microwave connections? Does it take EDGA four milliseconds or more to report a trade? Of course not. Competitive timings these days for co-located firms at exchanges are lower than 100 *microseconds*, and often enough much lower.

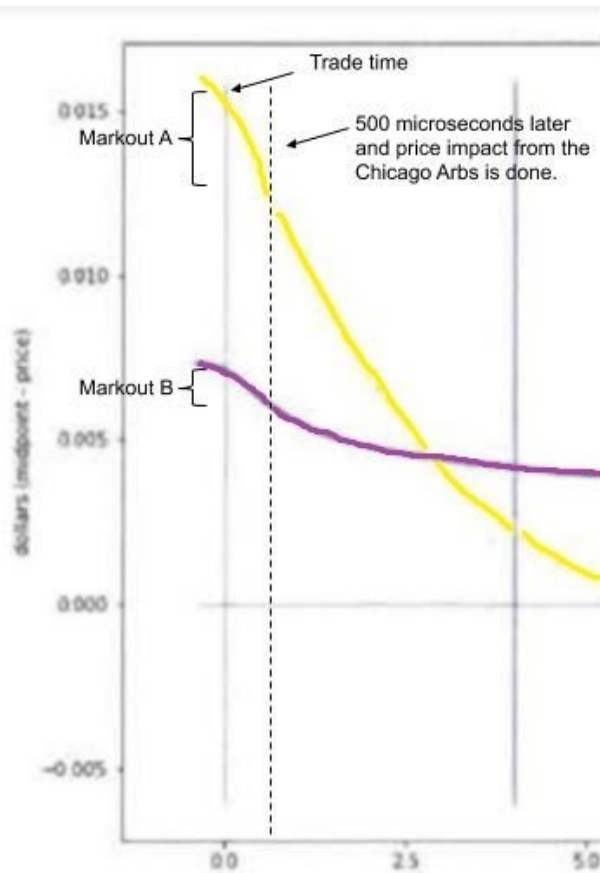


I don't see how the EDGA's core evidence for the Filing, the Yellow Sample, shows Chicago Arbs picking off stale quotes. If anything the trades where market makers did not attempt to cancel their quotes (because they know the order has already traded), the Purple Sample, includes Chicago Arb activity. The Cboe Letter describes the Purple Sample as showing "relatively stable prices following an execution."

Another possibility is that the Yellow Sample actually demonstrates that EDGA market makers *already* use microwave between Chicago and New Jersey, suffering occasional missed cancels for whatever reason, or that the market maker's signal to cancel originates much closer to New Jersey than EDGA has described, with serious implications for the duration of any proposed speed bump cure. It could also be that all this evidence is nonsense.

*No material harm*

Moving ahead anyway and stipulating to the Cboe Letter's presentation, let's deconstruct the SPY graph. The main area of interest is on the left of the graph, in the first few moments after trade time. If the Chicago Arbs get from Chicago to New Jersey in about four milliseconds, at trade time - that is,  $t_0$  in its markout calculations or at 0.0 milliseconds in the graph (the graph origin) - for the Yellow Sample, and assuming all those trades were initiated by Chicago Arbs, as noted above we're already four or more milliseconds along from the relevant event in Chicago. In that time the Chicago Arbs have shipped their data to New Jersey and are picking off stale quotes wherever they can (hence the trade itself in the Yellow Sample, of course).



The relevant analysis period to understand the price impact of the Chicago Arbs, then, is in the first few hundred microseconds after the trade at  $t_0$  and not any further along. That's because by the trade time at  $t_0$  the Chicago Arbs as a group have received their data from Chicago, processed it, and executed their trades. All that's left is to generate a new NBBO.<sup>22</sup> We'll generously add 500 microseconds past  $t_0$  for EDGA and other exchanges to report trades and new quotes to the SIP and for the SIP to process them and update the NBBO (the NBBO determines the price impact shown in the graph). After those 500 microseconds, *all price impact from the Chicago Arbs is long since finished regardless of when market makers try to cancel their quotes.*

The markout attributable to the Chicago Arbs - and that's assuming there aren't any coincidences and all the activity in the Yellow Sample is due to arbitrage<sup>23</sup> - is about .25 cents ("Markout A" in the graph). The control sample, the Purple Sample, has a markout about half as large, or about .125 cents ("Markout B" in the graph). In other words, taking the graph as presented and under the most favorable assumptions to the Filing, the "harm" done on EDGA by the Chicago Arbs is (Markout A - Markout B), or about .125 cents per share.

<sup>22</sup> If the Yellow Sample includes all trades where a market maker tried but failed to cancel its quote before a trade was executed against it, following EDGA's narrative all successful arbitrage behavior from the Chicago Arbs as a group must then be included in the graph at  $t_0$ .

<sup>23</sup>Without supporting evidence the Cboe Letter suggests that all the trades included in the Yellow Sample are done by Chicago Arbs and all the trades in the Purple Sample are done by "investors." If I understand Cboe, in the entire month of July 2019 there aren't any coincidental failed cancels in the Yellow Sample, there aren't any investors taking liquidity in the Yellow Sample, and there aren't any Chicago Arbs taking liquidity in the Purple Sample.

What's that work out to? We'll have to estimate that because EDGA hasn't provided any data. Looking at 2019 trades on EDGA in SPY, it could be as high as \$90 a day.<sup>24</sup> EDGA wants to make one of the most significant changes to U.S. market structure in years because the Chicago Arbs are costing its market makers less than \$100 a day, if anything at all.

.....

*6. Is the proposal tailored in a manner such that its potential benefits outweigh the potential or likelihood of harm or unintended consequences to the national market system?*

No.

Commenters have debated what research on TSX Alpha, a Canadian exchange, shows or doesn't show about discriminatory speed bumps at an inverted fee exchange. An academic study is pretty gloomy about it.<sup>25</sup> An IIROC/Bank of Canada study<sup>26</sup> is less gloomy, though hardly enthusiastic, and that's the study the Cboe Letter highlights.<sup>27</sup> The Cboe Letter includes a quotation from an IIROC notice about the study, that TSX Alpha "did not adversely affect the quality of Canadian markets,"<sup>28</sup> but that's pretty weak beer if anyone is looking for any kind of endorsement. If someone asked me about a new business partner and my answer was "Well, they didn't adversely affect the quality of life," I don't think that would impress.

Contrary to EDGA's pitch that its own speed bump "will improve market quality for investors"<sup>29</sup> and "make better markets"<sup>30</sup> and "promote liquidity provision"<sup>31</sup> and "improve displayed prices"<sup>32</sup> and "benefit all market participants,"<sup>33</sup> the IIROC/Bank of Canada study found TSX Alpha "did not impact market-wide liquidity" and could not "identify any significant impacts on effective spreads, price impact or quoted depth."<sup>34</sup> For certain participants, though, it found negative effects. Buy-side investors experienced "higher price impacts and effective spreads..."<sup>35</sup>

The Ontario Securities Commission ("OSC") staff notice about the study echoes IIROC/Bank of Canada and goes further. OSC reported that its own "market quality measures examined did not materially change..."<sup>36</sup> and that in OSC's survey of market participants, TSX Alpha "added complexity into routing decisions" and "in certain situations, fill rates on Alpha have decreased, often for orders that are expected to go through multiple price levels or need to be split and sent to multiple marketplaces simultaneously (e.g. institutional orders). Some dealers reported initial fill rates to be much lower on Alpha in these circumstances..."<sup>37</sup> None of which seems like any kind of an endorsement at all, and to my ear sounds like quote fading ahead of institutional investor equity market sweeps, making displayed prices less accessible.

---

<sup>24</sup> Based on back-of-the-envelope calculations for Q2 2019 SPY trade data on EDGA. Assumptions here include that the Yellow Sample is 5% of all EDGA trade volume and a .125 cents per share markout attributable to Chicago Arbs, as above. If EDGA doesn't like this estimate, it can provide its own.

<sup>25</sup> See note 7.

<sup>26</sup> Anderson, Andrews, Devani, Mueller, Walton, "Speed Segmentation on Exchanges: Competition for Slow Flow," Bank of Canada Staff Working Paper 2018-3, January, 2018. ("IIROC/BC")

<sup>27</sup> Cboe Letter, pages 10-11.

<sup>28</sup> Cboe Letter, pages 10-11.

<sup>29</sup> Cboe Letter, page 17.

<sup>30</sup> Filing, page 2.

<sup>31</sup> Filing, page 4.

<sup>32</sup> Filing, page 33.

<sup>33</sup> Filing, page 33.

<sup>34</sup> IIROC/BC, page 16.

<sup>35</sup> IIROC/BC, page 16.

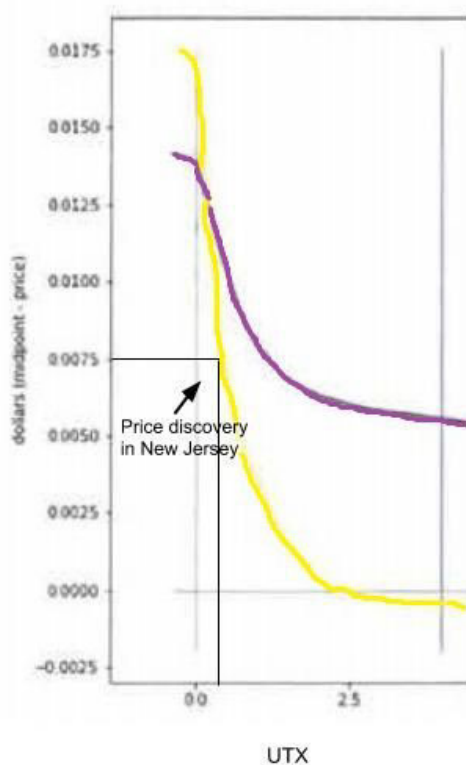
<sup>36</sup> OSC Notice, page 3.

<sup>37</sup> OSC Notice, page 4.

### *The rent is too damn high*

Though EDGA justifies its proposal because of the Chicago Arbs and a handful of ETPs like SPY, I believe the real impact of the Filing will be found nowhere near Chicago. The real action will be in the National Market System data centers in New Jersey, with every security traded in those data centers, and with obvious regulatory rent-seeking.

There's a good sense of all this in the graphs Cboe provides in the appendix to its letter.<sup>38</sup> The graph for UTX is instructive. In that example the price in the Yellow Sample drops like a stone in the first 200 microseconds or so, realizing approximately half of its overall price movement, moving much more quickly than it does in the SPY example.



Are the Chicago Arbs arbitraging UTX against the S&P futures contracts, trading against the equities markets in UTX much more rapidly than they do the markets in SPY? Of course not. The UTX graph is likely showing the effect of investor equities market sweeps and has little or nothing to do with the Chicago futures markets.

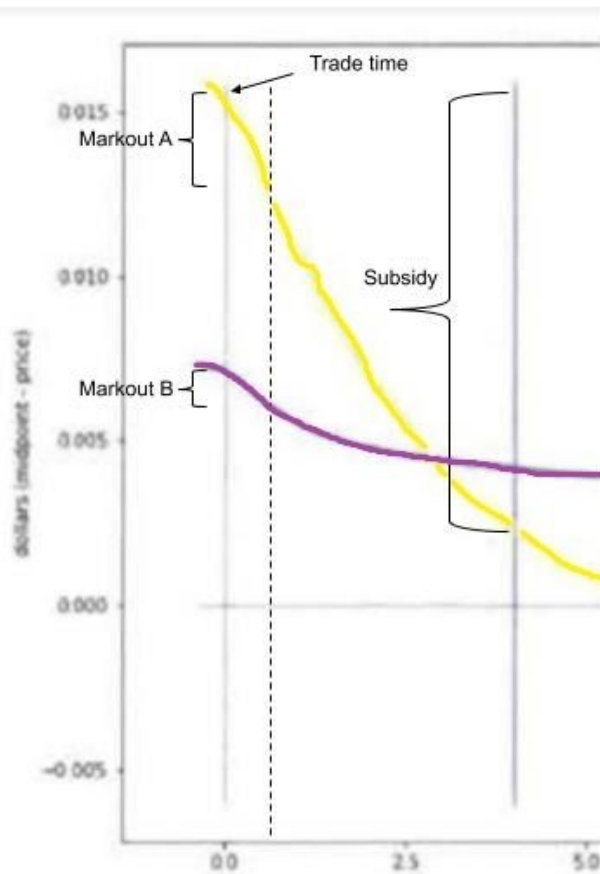
Since all the major equities market data centers are clustered in northern New Jersey, and since price discovery for a corporate stock is in the equities markets themselves, the price impact of an equities market sweep will unfold very quickly after  $t_0$ . The Chicago Arb story doesn't scale for corporate stocks, but EDGA's proposed speed bump will be the same in UTX as it is in SPY.

If the speed bump is implemented, I've argued that EDGA's market makers will receive a valuable regulatory subsidy.<sup>39</sup> We can use the graph to estimate what that investor-funded subsidy amounts to for trading in SPY.

<sup>38</sup> Cboe Letter, pages 18-20.

<sup>39</sup> Letter from R. T. Leuchtkafer to Vanessa Countryman, Secretary, SEC, July 12, 2019 ("Leuchtkafer 2").





With a four millisecond speed bump that investor-funded subsidy could be \$900 a day or more in SPY.<sup>40</sup> If we extrapolate that to all stocks, the total daily subsidy on EDGA is \$75,000.<sup>41</sup> Extrapolating that to the market at large, conservatively, it works out to about \$1 billion a year.<sup>42</sup> That subsidy will come from investors attempting but failing to trade with advertised prices and paying worse prices as a result.

.....

*7. Should the Exchange's unprotected, manual quote be allowed to lock or cross manual quotations disseminated by another manual market? Why or why not?*

No. Reg NMS defines a manual quote as any quotation which is not "immediately and automatically accessible"<sup>43</sup> and a protected quote as a quotation which is immediately and automatically accessible. As I argued in the IEX approval process, the intent behind a "protected quote" wasn't to guarantee

<sup>40</sup>Assumptions here include that the Yellow Sample is 5% of all EDGA SPY trade volume and a penny per share price impact avoided.

<sup>41</sup> Based on these simple variables: 150 million shares/day traded, 5% sweep/arb fade rate, a penny per share price impact avoided.

<sup>42</sup> Taking \$75,000 a day for EDGA and then estimating for the market based on EDGA's approximate 2% market share of NMS stocks.

<sup>43</sup> Reg NMS adopting release, Exchange Act Release No. 3451808, June 9, 2005, page 1. ("Adopting Release")

instantaneous execution but to firewall human intervention ahead of execution.<sup>44</sup> The SEC accepted this interpretation.<sup>45</sup>

Under the Filing, EDGA's manual quotations guarantee the prospect of human intervention *after* an order has been received by the exchange but *before* that order is executed. That intervention is encoded in participant algorithms, which makes it faster but no less of an intervention than if a guy named Tommy did it. Discretion isn't somehow scrubbed away because it's contained in software. Here's an example:

- A. TOMMY: If we ain't good, I'm backin' out.
- B. ALGORITHM: if (sux == mktcond()) backout();

To justify exemptive relief for its manual quotations, EDGA needs a convincing argument of how the discretion embedded in the software of (B) is materially different from the discretion embedded in the "wetware" of (A). It also needs to explain why delaying marketable orders on receipt to give participants time to exercise discretion isn't exactly what the Adopting Release contemplated when it defined a manual quote. Crucially, it doesn't matter that participants won't see marketable orders in EDGA's queue because they'll often enough see their brethren executing on other markets. More crucially, in defining a manual quote, the Adopting Release doesn't stipulate participants on an exchange have to first see an arriving order, but only that they can exercise discretion at any time *after* an order has arrived at the exchange and *before* it is executed:

First and most importantly, Rule 611 protects only immediately accessible quotations that are available through automatic execution. It does not require investors submitting marketable orders to access "maybe" quotations that, *after arrival of the order, are subject to human intervention and thereby create the potential for other market participants to determine whether to honor the quotation.*<sup>46</sup> [Emphasis added.]

The language is broad and quite clear. EDGA's fanciful "otherwise automated market"<sup>47</sup> is a fiction. It's not a new marketplace category, it's a silicon-powered version of the old-fashioned specialist market, though with fewer constraints on specialist behavior. EDGA will allow its market makers plenty of opportunity to advertise quotes but then back away as they wish, how they wish, and when they wish. Our friend Tommy would have *loved* it. If this is how EDGA now hopes to run its business all it's entitled to is a simple manual quote without any privileged status.<sup>48</sup>

.....

**8. What impact, if any, would the dissemination of an unprotected, manual quote have on the national market system? Should EDGA's unprotected, manual quote be disseminated by the SIP? If so, should the SIP disseminate a modifier to indicate that EDGA's quote is manual? Should the EDGA quote be used to calculate the NBBO? Should the EDGA quote be used to calculate midpoint values?**

---

<sup>44</sup> Letter from R. T. Leuchtkofer to Brent J. Fields, Secretary, SEC, February 19, 2016. ("Leuchtkofer 3").

<sup>45</sup> "Notice of Proposed Commission Interpretation Regarding Automated Quotations Under Regulation NMS," Release 34-77047, March 18, 2016.

<sup>46</sup> Adopting Release, page 119.

<sup>47</sup> Filing, pages 13-14.

<sup>48</sup> To be clear about it, I am not arguing a marketable order in a speed bump queue has been "presented" to a market maker, as with Rule 602. Under the Adopting Release and for the purposes of determining whether an exchange is entitled to publish anything more than a conventional manual quote, however, an order has "arrived" at an exchange if (a) it is held in an exchange's speed bump queue while (b) other market participants can still determine whether to honor their quotations before that order is released for execution.

I agree with the superb commentary by Tyler Gellasch of Healthy Markets on these points.<sup>49</sup>

.....

*9. How will the dissemination of EDGA's unprotected, manual quote impact a broker dealer's obligation to obtain best execution?*

As with question eight, I agree with the superb commentary by Tyler Gellasch of Healthy Markets on this point.<sup>50</sup>

.....

*10. What would be the impact, if any, on the national market system if other national securities exchanges, with a larger percentage of overall trading volume, adopted a similar proposal? In particular, how would the proposal affect market quality?*

and

*11. What are commenters' views on how the proposal would affect trading activity, in general, and liquidity providers, in particular, on other markets? Would the LP2 delay mechanism impose systemic risks and create informational disparities across the national market system? Would the proposal provide EDGA liquidity providers with the option to leverage or free ride price discovery that occurs at other trading venues?*

A discriminatory speed bump implemented within a distributed, fragmented marketplace is just simple gamesmanship. The ills it promises to cure can exist - if they exist - only so long as any other market remains unbiased. If all other markets implement a discriminatory speed bump of the same duration, absolutely nothing has been accomplished. If other markets adopt speed bumps with different durations, that entire marketplace will stumble deep into a nightmare of discriminatory speed bump, smart router, and network latency gaming. Unlike IEX's universal speed bump, which compensates for a limited geography, applies universally, and simply allows the exchange to maintain accurate pricing for an undisplayed and unpriced order type, EDGA's proposed discriminatory speed bump empowers privileged intermediaries to extract rents from participants on every unbiased marketplace in the National Market System. Those rents will come from free-riding on the price discovery of the unbiased markets and from using disparate latencies to spot equity market sweeps and run from them.

It often seems intermediaries have come to believe their only responsibility today is to collect spreads on reasonably balanced uninformed order flow and that it's the SEC's job to help them. Another way of stating this is they've come to believe their only responsibility is to make risk-free profits. That delusion would be fine if, as in some markets, they weren't given a variety of regulatory subsidies to make markets - subsidies like rule exemptions, relaxed capital requirements, preferred pricing, various order book priorities, and more. But they do have those regulatory subsidies, and to earn them they are expected to consistently make markets, and, crucially, to make markets even at the risk of trading with informed order flow.

*Oversize me*

As markets evolved in the last 20 years or so, the market maker's business model came to include regularly oversizing its aggregate trading interest across dispersed marketplaces. A market maker would put up quotes on as many as a dozen different markets and when it was hit on one it would race to cancel its other quotes. It took a while but investors eventually understood that market makers had absolutely no intention of trading all their advertised size at once. This fact was a central plotline in Michael Lewis's

---

<sup>49</sup> Letter from Tyler Gellasch, Executive Director, Healthy Markets Association to Vanessa Countryman, Secretary, SEC, July 16, 2019 ("Healthy Markets Letter").

<sup>50</sup> Healthy Markets Letter.

*Flash Boys* and a sore spot for anyone who still believed orders should be *bona fide*. It is an obvious species of spoofing with real market harm, but regulators continue to tolerate it. I addressed this in a series of comment letters about a Bats rule filing:<sup>51</sup>

HFT market makers portray themselves as hapless victims of adverse selection, but it's only because they want to display size on multiple exchanges that they're at risk. There's nothing that compels these firms to post any particular size orders on eight or even ten exchanges at once. They do it because they've found it more profitable to post size on multiple exchanges, and then cancel and reprice when they're hit, than it is to post smaller size on those exchanges and honor their quotes. A firm might want to trade no more than 1,000 shares at a time, but it doesn't know where it will find a contra, and its eyes are bigger than its stomach. So it posts that 1,000 shares on as many as ten exchanges, making 10,000 shares at risk though it only wants to trade 1,000. If the firm's hit at any one of the them, it cancels the other quotes and reprices.

Nothing compels it to post 1,000 shares in ten different places. It could just as easily post 100 shares on each of those ten exchanges. It could post 1,000 shares on only one exchange. It could post 500 shares on each of two exchanges. It has any number of options but the one it chooses is likely its most profitable, and what so many firms choose to do is to post more liquidity than they are in fact willing to trade, planning beforehand to cancel everything they can if a contra appears, all in the hope of maximizing profits. It is a very deliberate strategy that misleads the market because it is by now long past obvious the aggregate displayed size in today's HFT marketplace is "not representative of actual supply and demand."

For market participants, this behavior is also profoundly anticompetitive. If all a firm wants to trade is 1,000 shares at a price, once again, the firm could post 100 shares on each of ten exchanges. If it posts only 100 shares at a market, though, a contra at that market is much more likely to exhaust the firm's quote and trade with its competitors. It could also post 1,000 shares on only one exchange, or 500 shares each on two exchanges. If the firm isn't on all exchanges at once, however, it might miss a contra, and that contra will trade with its competitors. So by posting the full 1,000 on every exchange the firm takes the biggest bite it can stomach and then rushes to fade everywhere else. Since fading successfully depends on speed, firms spend whatever it takes to be fast. The effect is that smaller competitors on any given market are squeezed out in the spoofing game, and the marketplace is soon dominated by a handful of large HFT firms with the resources to win the speed arms race.

As for market centers, on the other hand, we can wonder whether this behavior helps sustain market fragmentation. If HFT market makers posted no more than their *bona fide* interest and cut back where they traded as a result, perhaps we wouldn't have so many lit markets, and perhaps exchange groups like BATS would finally consolidate their order books. BATS has lectured the SEC why it believes Reg NMS contributes to market complexity, but BATS should consider instead how tolerating this kind of HFT behavior contributes to market complexity.<sup>52</sup>

Thanks to *Flash Boys* investors have wised up to this game and now regularly employ smart order routers to synchronize and sweep markets in search of liquidity. Critics were scolded for years that ever speedier markets impounded new information into prices more efficiently, with spectacular benefits for all investors.

---

<sup>51</sup> See for example letter from R. T. Leuchtkafer to Brent J. Fields, Secretary, SEC, September 4, 2015 available at <https://www.sec.gov/comments/sr-bats-2015-57/bats201557-3.pdf> ("Leuchtkafer 4").

<sup>52</sup> Leuchtkafer 4, page 5.

But when investors caught on to the game, suddenly the story became speed kills and the SEC is now approached for relief. No one would say "Thanks to *Flash Boys* many institutional investors now use fast smart routers, and we want a discriminatory speed bump so our high frequency traders can pull quotes in front of an investor's sweep in the dispersed price/time equities marketplace we designed, built, or bought." They won't say so because the SEC's explicit policy goal for nearly a half-century has been to make displayed quotes *more* accessible to investors. But what else do EDGA's graphs show? The story about fat money arbitrageurs is just bait for regulatory rent-seeking.

That story collapses at a glance. The SEC is being asked to make displayed liquidity *less* accessible, to make it more difficult for investors to trade with displayed prices. That's what a discriminatory speed bump does, by design. Now that everyone understands how things work, if market makers can't figure out how to trade in the marketplace they helped create, they should make room for firms that can.

In the meantime investor equity market sweeps are the only discipline holding market makers back from oversizing - and then fading - even more than they do today. The SEC should not make that game easier.

Sincerely,

R. T. Leuchtkafer