

Flash Boys: A Wall Street Revolt

May 2014

"Flash Boys: A Wall Street Revolt," authored by Michael Lewis, tells the story of how the U.S. financial markets that were once stable and predictable have become predatory by feeding on the ordinary investor and how a group of dedicated individuals set out to correct the situation. 

Brad Katsuyama was a trader in equities for the Royal Bank of Canada (RBC). In 2002, he was transferred by RBC to work in its New York City trading office.

Brad's main role as a trader was to provide liquidity to the market. He sat between buyers and sellers who wanted to trade large amounts of stock and the public markets where the volumes were smaller. For instance, if a large pension plan wanted to sell three million shares of Intel, but the market showed a demand for only one million shares, he would buy one million shares from the pension plan and immediately resell it in the market. He would then work artfully over the next few hours to unload the other two million shares.

In the seven years of his trading career, he had trusted his trading terminals. If his trading screens showed 10,000 shares of Intel offered at \$22 per share, he could buy 10,000 shares of Intel at \$22 per share. In 2006, RBC decided to move to an electronic trading system and paid \$100 million for a U.S. electronic stock trading firm called Carlin Financial. Shortly after being forced to use Carlin's technology, Brad encountered severe problems in trading. His trading terminals were no longer reliable. If they showed a market of 10,000 shares of Intel offered at \$22, and he pushed the Buy button, the offers vanished. He could buy perhaps 100 shares at \$22, and then the price moved inexplicably up. It was as if the market had read his mind. He found that his fellow traders were experiencing the same problem

He gathered the company's technologists around him and explained the problem. Their response was that it was Brad's fault, or it was because the market was too active. Brad ran an experiment to prove his point. His trading terminal showed a total of 100,000 shares of AMD offered at \$48 distributed between the New York Stock Exchange, Nasdaq, the BATS electronic exchange, and the Direct-Edge electronic exchange. He entered an order into his terminal to buy 100,000 shares of AMD at \$48 per share. He watched for several seconds, and the market didn't move. He then hit the Enter key, and the offerings all disappeared and were replaced with higher offers. Someone was reading his mind, or so it seemed.

Brad set out to build a team of people to investigate the cause of this and other anomalies in the stock markets. The team designed a series of experiments, not to make money but to test theories. to answer the question: Why was there a difference between the stock market displayed on his trading terminals and the actual market? Why, when he went to buy 20,000 shares of Intel offered on his trading screens, did the market only sell him 2,000 shares?

At the time, there were thirteen different stock exchanges operating in the New York City/New Jersey area. A major clue came when the team found that if an order was routed to only one exchange, the team was able to buy everything offered by the exchange. Furthermore, as they increased the number of

exchanges, the percentage of the order that was filled decreased. The more places from which they tried to buy stock, the less they actually bought.

The team then tried a new experiment. They introduced delays into their order streams so that all orders arrived at all exchanges at substantially the same time. Surprise! The orders were all executed at the markets offered by the exchanges.

It now became clear that certain traders were taking advantage of the team's order arriving at one exchange to beat that order to other exchanges that were further away time-wise. At these exchanges, they would sell a small amount of the stock, thus driving the price down. When the team's buy order appeared, they would buy at the reduced price. The difference in time was measured only in milliseconds, but it was enough that a firm with an extremely fast trading engine could take advantage of. By this time, there were many firms that had this capability – they were the new high-frequency trading firms (HFTs).

HFTs use special algorithms and extremely fast servers collocated with the exchange matching engines to gain a speed advantage over other brokers. The practice that Brad and his team had discovered was called *front-running*. An HFT would post small 100-share buy/sell orders for many stocks on multiple exchanges. These orders were not meant to make money – they were intended to detect what other brokers were doing. As soon as one of these test orders was executed, the HFT had a good idea that a much larger block was behind the trade. They would then, at very high speeds, execute against that order on the other exchanges so that the original investor was faced with a market that had suddenly moved against him. Though the difference in stock price might just be a penny or two, later calculations showed that this amounted to billions of dollars a year taken out of investors' pockets.

Brad and his team devised a trading engine that attempted to equalize the time at which trades arrived at various exchanges. Named Thor, the trading engine introduced delays in the paths to the various exchanges so that the time for an order to travel to any exchange was the same. Thor was marketed by RBC with some success. However, it was plagued by differences in propagation time over which it had no control. The nominal transit time for an order to some specific exchange might be ten milliseconds. However, this time delay was subject to a great deal of variance depending upon which route the common carrier was using at the time and the traffic load on the channels. The time delay could vary by several milliseconds, thwarting Thor's efforts to reach each exchange at the same time.

Further tests showed that the minimum time that it took to travel over a channel to an exchange was about 450 microseconds. Therefore, the time difference between the arrivals of an order at different exchanges had to be less than 450 microseconds. This could only be accomplished by running private fiber channels from the firm's trading engine to the various exchanges. It was determined that a 350-microsecond window would be sufficient to guarantee fair trading by eliminating front-running.

In January, 2012, Brad and several members of his team left RBC to build their own exchange based on this insight. Named the Investor's Exchange, they quickly decided to change the name when they saw their URL – www.investorsexchange.com. The name of their new exchange became IEX. A core part of the team was the "Puzzle Masters." They were tasked to find every way to break the exchange's fairness policies and to determine defenses against these methods.

When trying to raise capital to start IEX, the team discovered that there was a lot of resistance to the concept, especially among the large Wall Street banks. The banks offered preferential rates to HFTs to handle their orders and made significant income from front-running and other predatory practices that IEX sought to eliminate. However, the capital was raised, and IEX laid its private fiber network and built its own matching engine.

In addition to ensuring that market pricing data arrived at external points-of-presence simultaneously, IEX also instituted policies that avoided other predatory practices, including:

- Following a published set of rules in the exchange's computerized order matching engine. The rules for order matching followed by many exchanges and large Wall Street banks were kept as proprietary information.
- Offering a limited number of simple and familiar order types. There were only four – market, limit, mid-point peg, and fill or kill. Over 150 order types existed at other exchanges, all meant to give some firm an advantage.
- Charging a fixed fee of 0,09 cents per share on most orders (or a flat percentage rate of 0.3% on stocks worth less than \$1 per share). Other exchanges charged a wide range of fees or even paid for order flow to attract the type of order traffic they wanted. IEX refused to pay for order flow.
- Slightly delaying market pricing data to all customers so that no significant advantage can be gained by a broker locating its computers in close proximity to the IEX exchange. Colocation of a broker with the IEX systems was prohibited.
- Limiting ownership to a group of "ordinary (non HFT) investors" who are required to route orders through other brokers (who are not allowed to be investors in the company), which alleviates many conflict-of-interest issues.

IEX went into operation in October, 2013. During that month, it traded about 2 million shares. In April, 2014, its trading volume had skyrocketed to 57 million shares. This is still small compared to the big exchanges, but IEX's growth is on a path to make it an important exchange in the financial marketplace. The key to its success is to get the big banks to buy in to the concept of a fair exchange. So far, Goldman Sachs has taken the plunge. It remains to be seen if IEX's efforts to create a fair marketplace will succeed or if greed will triumph.

This description has highlighted only one aspect of the predatory practices in today's financial markets – front-running. However, "Flash Boys" exposes a myriad of other practices, their impact on the normal investor, and what IEX is doing to hopefully curb these practices. It is a comprehensive exposé of what predatory practices today's financial markets have subtly adopted, practices that are not understood by even sophisticated investors.