Notes and highlights for

flash boys a wall street revolt_lewis, michael (1)

Lewis, Michael

Chapter 3: Ronan's Problem

Highlight (yellow) - Location 772

That was another thing Ronan learned: A lot of people in and around the telecom industry were more knowing than knowledgeable.

Highlight (yellow) - Location 816

It was only a matter of time before the stock exchanges figured out that, if people were willing to spend hundreds of thousands of dollars to move their machines around inside some remote data center just so they might be a tiny bit closer to the stock exchange, they'd pay millions to be inside the stock exchange itself. Ronan followed them there. He came up with an idea: sell proximity to Wall Street as a service. Call it "proximity services." "We tried to trademark proximity, but you can't because it's a word," he said. What he wanted to call proximity soon became known as "co-location," and Ronan became the world's authority on the subject. When they ran out of ways to reduce the length of their cable, they began to focus on the devices on either end of the cable. Data switches, for instance. The difference between fast data switches and slow ones was measured in microseconds (millionths of a second), but microseconds were now critical. "One guy says to me, 'It doesn't matter if I'm one second slower or one microsecond; either way I come in second place.' " The switching times fell from 150 microseconds to 1.2 microseconds per trade. "And then," says Ronan, "they started to ask, 'What kind of glass are you using?' " All optical fibers were not created equal; some kinds of glass conveyed light signals more efficiently than others. And Ronan thought: Never before in human history have people gone to so much trouble and spent so much money to gain so little speed. "People were measuring the length of their cables to the foot inside the exchanges. People were buying these servers and chucking them out six months later. For microseconds."

Highlight (yellow) - Location 839

By the end of 2007 Ronan was making hundreds of thousands of dollars a year building systems to make stock market trades faster. He was struck, over and over again, by how little the traders he helped understood of the technology they were using. "They'd say, 'Aha! I saw it— it's so fast!' And I'd say, 'Look, I'm happy you like our product. But there's no fucking way you saw anything.' And they're like, 'I saw it!' And I'm like, 'It's three milliseconds— it's fifty times faster than the blink of an eye.' " He was also keenly aware that he had only the faintest idea of the reason for this

incredible new lust for speed. He heard a lot of loose talk about “arbitrage,” but what, exactly, was being arbitraged, and why did it need to be done so fast? “I felt like the getaway driver,” he said. “Each time, it was like, ‘Drive faster! Drive faster!’ Then it was like, ‘Get rid of the airbags!’ Then it was, ‘Get rid of the fucking seats!’ Towards the end I’m like, ‘Excuse me, sirs, but what are you doing in the bank?’ ” He had a sense of the technological aptitude of the various players. The two biggest high-frequency trading firms, Citadel and Getco, were easily the smartest. Some of the prop shops were smart, too. The big banks, at least for now, were all slow.

Highlight (yellow) - Location 957

The funny thing was that a lot of what Ronan had seen and heard didn’t make sense to him: He didn’t know what he knew. Brad now helped him to understand. For instance, Ronan had noticed the HFT guys creating elaborate tables of the time, measured in microseconds, it took for a stock market order to travel from any given brokerage house to each of the exchanges. “Latency tables,” these were called. The times were subtly different for every brokerage house— they depended upon where the brokerage house physically was located and which fiber networks it leased in New Jersey. These tables took trouble to create and were of obvious value to high-frequency traders, but Ronan had no idea why. This was the first Brad had heard of latency tables, but he knew exactly why they had been created: They enabled high-frequency traders to identify brokers by the time their orders took to travel from one exchange to the other. Once you had figured out which broker was behind any given stock market order, you could discern patterns in each broker’s behavior. If you knew which broker had just come into the market with an order to buy 1,000 shares of Intel, you might further guess whether those 1,000 shares were the entire order or a part of a much larger order. You might also guess how the broker might distribute the order among the various exchanges and how much above the current market price for Intel shares the broker might be willing to pay. The HFT guys didn’t need perfect information to make riskless profits; they only needed to skew the odds systematically in their favor. But, as Brad put it, “What you’re looking for ultimately is large brokers who are behaving idiotically with their customers’ orders. That’s the real gold mine.”

Chapter 4: Tracking the Predator

Highlight (yellow) - Location 1208

Schwall wanted to think of himself as a guy who lived by a few simple principles, a good soldier. After the financial crisis he was more like the Resentful Butler. He had a taste for asking complicated questions, and for tracking the answers into whatever rabbit hole they might lead him. He had, in short, an obsessive streak.

Highlight (pink) - Location 1315

Several days later he’d worked his way back to the late 1800s. The entire history of Wall Street was the story of scandals, it now seemed to him, linked together tail to trunk like circus elephants. Every systemic market injustice arose from some loophole in a regulation created to correct some prior injustice. “No matter what the regulators did, some other intermediary found a way to react, so there would be another form of front-running,” he said.

Highlight (pink) - Location 1386

The argument in favor of high-frequency traders had beaten the argument against them to the U.S. regulators. It ran as follows: Natural investors in stocks, the people who supply capital to companies, can't find each other. The buyers and sellers of any given stock don't show up in the market at the same time, so they needed an intermediary to bridge the gap, to buy from the seller and to sell to the buyer. The fully computerized market moved too fast for a human to intercede in it, and so the high-frequency traders had stepped in to do the job. Their importance could be inferred from their activity: In 2005 a quarter of all trades in the public stock markets were made by HFT firms; by 2008 that number had risen to 65 percent. Their new market dominance— so the argument went— was a sign of progress, not just necessary but good for investors. Back when human beings sat in the middle of the stock market, the spreads between the bids and the offers of any given stock were a sixteenth of a percentage point. Now that computers did the job, the spread, at least in the more actively traded stocks, was typically a penny, or one-hundredth of 1 percent. That, said the supporters of high-frequency trading, was evidence that more HFT meant more liquidity. The arguments against the high-frequency traders hadn't spread nearly so quickly— at any rate, Brad didn't hear them from the SEC. A distinction cried out to be made, between "trading activity" and "liquidity." A new trader could leap into a market and trade frantically inside it without adding anything of value to it. Imagine, for instance, that someone passed a rule, in the U.S. stock market as it is currently configured, that required every stock market trade to be front-run by a firm called Scalpers Inc. Under this rule, each time you went to buy 1,000 shares of Microsoft, Scalpers Inc. would be informed, whereupon it would set off to buy 1,000 shares of Microsoft offered in the market and, without taking the risk of owning the stock for even an instant, sell it to you at a higher price. Scalpers Inc. is prohibited from taking the slightest market risk; when it buys, it has the seller firmly in hand; when it sells, it has the buyer in hand; and at the end of every trading day, it will have no position at all in the stock market. Scalpers Inc. trades for the sole purpose of interfering with trading that would have happened without it. In buying from every seller and selling to every buyer, it winds up: a) doubling the trades in the marketplace and b) being exactly 50 percent of that booming volume. It adds nothing to the market but at the same time might be mistaken for the central player in that market. This state of affairs, as it happens, resembles the United States stock market after the passage of Reg NMS. From 2006 to 2008, high-frequency traders' share of total U.S. stock market trading doubled, from 26 percent to 52 percent— and it has never fallen below 50 percent since then. The total number of trades made in the stock market also spiked dramatically, from roughly 10 million per day in 2006 to just over 20 million per day in 2009. "Liquidity" was one of those words Wall Street people threw around when they wanted the conversation to end, and for brains to go dead, and for all questioning to cease. A lot of people used it as a synonym for "activity" or "volume of trading," but it obviously needed to mean more than that, as activity could be manufactured in a market simply by adding more front-runners to it. To get at a useful understanding of liquidity and the likely effects of high-frequency trading on it, one might better begin by studying the effect on investors' willingness to trade once they sense that they are being front-run by this new front-running entity. Brad himself had felt the effect: When the market as displayed on his screens became illusory, he became less willing to take risk in that market— to provide liquidity. He could only assume that every other risk-taking intermediary— every other useful market participant— must have felt exactly the same way. The argument for HFT was that it provided liquidity, but what did this mean? "HFT firms go home flat every night," said Brad. "They don't take positions. They are bridging an amount of time between buyers and sellers that's so small that no one even knows it exists." After the market was computerized and decimalized, in 2000, spreads in the market had narrowed— that much was true. Part of that narrowing would have

happened anyway, with the automation of the stock market, which made it easier to trade stocks priced in decimals rather than in fractions. Part of that narrowing was an illusion: What appeared to be the spread was not actually the spread. The minute you went to buy or sell at the stated market price, the price moved. What Scalpers Inc. did was to hide an entirely new sort of activity behind the mask of an old mental model— in which the guy who "makes markets" is necessarily taking market risk and providing "liquidity." But Scalpers Inc. took no market risk.§ In spirit Scalpers Inc. was less a market enabler than a weird sort of market burden. Financial intermediation is a tax on capital; it's the toll paid by both the people who have it and the people who put it to productive use. Reduce the tax and the rest of the economy benefits. Technology should have led to a reduction in this tax; the ability of investors to find each other without the help of some human broker might have eliminated the tax altogether. Instead this new beast rose up in the middle of the market and the tax increased— by billions of dollars. Or had it? To measure the cost to the economy of Scalpers Inc., you needed to know how much money it made. That was not possible. The new intermediaries were too good at keeping their profits secret.¶ Secrecy might have been the signature trait of the entities who now sat at the middle of the stock market: You had to guess what they were making from what they spent to make it. Investors who eyeballed the situation did not find reason for hope. "There used to be this guy called Vinny who worked on the floor of the stock exchange," said one big investor who had observed the market for a long time. "After the markets closed Vinny would get into his Cadillac and drive out to his big house in Long Island. Now there is the guy called Vladimir who gets into his jet and flies to his estate in Aspen for the weekend. I used to worry a little about Vinny. Now I worry a lot about Vladimir." Apart from taking some large sum of money out of the market, and without taking risk or adding anything of use to that market, Scalpers Inc. had other, less intended consequences. Scalpers Inc. inserted itself into the middle of the stock market not just as an unnecessary middleman but as a middleman with incentives to introduce dysfunction into the stock market. Scalpers Inc. was incentivized, for instance, to make the market as volatile as possible. The value of its ability to buy Microsoft from you at $ 30 a share and to hold the shares for a few microseconds— knowing that, even if the Microsoft share price began to fall, it could turn around and sell the shares at $ 30.01— was determined by how likely it was that Microsoft's share price, in those magical microseconds, would rise in price. The more volatile Microsoft's share price, the higher Microsoft's stock price might move during those microseconds, and the more Scalpers Inc. would be able to scalp. One might argue that intermediaries have always profited from market volatility, but that is not really true. The old specialists on the New York Stock exchange, for instance, because they were somewhat obliged to buy in a falling market and to sell in a rising one, often found that their worst days were the most volatile days. They thrived in times of relative stability. Another incentive of Scalpers Inc. is to fragment the marketplace: The more sites at which the same stocks changed hands, the more opportunities to front-run investors from one site to another. The bosses at Scalpers Inc. would thus encourage new exchanges to open, and would also encourage them to place themselves at some distance from each other. Scalpers Inc. also had a very clear desire to maximize the difference between the speed of their private view of the market and the view afforded the wider public market. The more time that Scalpers Inc. could sit with some investor's stock market order, the greater the chance that the price might move in the interim. Thus an earnest employee of Scalpers Inc. would look for ways either to slow down the public's information or to speed up his own. The final new incentive introduced by Scalpers Inc. was perhaps the most bizarre. The easiest way for Scalpers Inc. to extract the information it needed to front-run other investors was to trade with them. At times it was possible to extract the necessary information without having to commit to a trade. That's what the "flash order" scandal had been

about: high-frequency traders being allowed by the exchanges to see other people's orders before anyone else, without any obligation to trade against them. But for the most part, if you wanted to find out what some big investor was about to do, you needed to do a little bit of it with him. For instance, to find out that, say, T. Rowe Price wanted to buy 5 million shares of Google Inc., you needed to sell some Google to T. Rowe Price. That initial market contact between any investor and Scalpers Inc. was like the bait in a trap— a loss leader. For Scalpers Inc., the goal was to spend as little as possible to acquire the necessary information— to make those initial trades, the bait, as small as possible. To an astonishing degree, since the implementation of Reg NMS, the U.S. financial markets had evolved to serve the narrow interests of Scalpers Inc. Since the mid-2000s, the average trade size in the U.S. stock market had plummeted, the markets had fragmented, and the gap in time between the public view of the markets and the view of high-frequency traders had widened. The rise of high-frequency trading had been accompanied also by a rise in stock market volatility— over and above the turmoil caused by the 2008 financial crisis. The price volatility within each trading day in the U.S. stock market between 2010 and 2013 was nearly 40 percent higher than the volatility between 2004 and 2006, for instance. There were days in 2011 in which volatility was higher than in the most volatile days of the dot-com bubble. The financial crisis brought with it a great deal of stock market volatility; perhaps people just assumed that there was supposed to be an unusual amount of drama in the stock market evermore. But then the financial crisis abated and the drama remained. There was no good explanation for this, but Brad now had a glimmer of one. It had to do with the way a front-runner operates. A front-runner sells you a hundred shares of some stock to discover that you are a buyer and then turns around and buys everything else in sight, causing the stock to pop higher (or the opposite, if you happen to be a seller). The Royal Bank of Canada had tested the effects on stock market volatility of using Thor, which stymied front-runners, rather than the standard order routers used by Wall Street, which did not. The sequential cost-effective router responded to the kickbacks and fees of the various exchanges and went to those exchanges first that paid them the most to do so. The spray router— which, as its name suggests, just sprayed the market and took whatever stock was available, or tried to— did not make any effort to compel a stock market order to arrive at the different exchanges simultaneously. Every router, when it bought stock, tended to drive the price of that stock a bit higher. But when the stock had settled— say, ten seconds later— it settled differently with each router. The sequential cost-effective router caused the share price to remain higher than the spray router did, and the spray router caused it to move higher than Thor did. "I have no scientific evidence," said Brad. "This is purely a theory. But with Thor the HFT firms are trying to cover their losses. I'm short when I don't want to be, so I need to buy to cover, quickly." The other two routers enabled HFT to front-run, so they wound up being long the stock. "[With] the other two, HFT is in a position to trade around a winning position," said Brad, "and they can do whatever they can do to force the stock even higher." (Or lower, if the investor who triggered the activity is a seller.) They had, in those privileged microseconds, the reckless abandon of gamblers playing with house money. The new choppiness in the public U.S. stock markets was spreading to other financial markets, as they, too, embraced high-frequency traders. It was what investors most noticed: They were less and less able to buy and sell big chunks of stock in a gulp. Their frustration with the public stock exchanges had led the big Wall Street banks to create private exchanges: dark pools. By the middle of 2011, roughly 30 percent of all stock market trades occurred off the public exchanges, most of them in dark pools. The appeal of these dark pools— said the Wall Street banks— was that investors could expose their big stock market orders without fear that those orders would be exploited.

Highlight (pink) - Location 1518
“I’m amazed that people don’t ask the questions,” he said. “That they don’t dig deeper. If some schmuck in West Chester, PA, can figure it out, I’ve got to believe other people did, too.”

Chapter 5: Putting a Face on HFT

Highlight (yellow) - Location 1861

He described the pleasure of his innovation this way: “It created something out of chaos. When you create something out of chaos, essentially, you reduce the entropy in the world.”

Chapter 6: How to Take Billions from Wall Street

Highlight (yellow) - Location 2194

Constantine was also Russian, born and raised in a small town on the Volga River. He had a theory about why so many Russians had wound up inside high-frequency trading. The old Soviet educational system channeled people away from the humanities and into math and science. The old Soviet culture also left its former citizens oddly prepared for Wall Street in the early twenty-first century. The Soviet-controlled economy was horrible and complicated but riddled with loopholes. Everything was scarce; everything was also gettable, if you knew how to get it. “We had this system for seventy years,” said Constantine. “People learn to work around the system. The more you cultivate a class of people who know how to work around the system, the more people you will have who know how to do it well. All of the Soviet Union for seventy years were people who are skilled at working around the system.” The population was thus well suited to exploit megatrends in both computers and the United States financial markets. After the fall of the Berlin Wall, a lot of Russians fled to the United States without a lot of English; one way to make a living without having to converse with the locals was to program their computers. “I know people who never programmed computers but when they get here they say they are computer programmers,” said Constantine. A Russian also tended to be quicker than most to see holes built into the U.S. stock exchanges, even if those holes were unintentional, because he had been raised by parents, in turn raised by their own parents, to game a flawed system.

Highlight (yellow) - Location 2246

they worked through the order types, they created a taxonomy of predatory behavior in the stock market. Broadly speaking, it appeared as if there were three activities that led to a vast amount of grotesquely unfair trading. The first they called “electronic front-running”— seeing an investor trying to do something in one place and racing him to the next. (What had happened to Brad, when he traded at RBC.) The second they called “rebate arbitrage”— using the new complexity to game the seizing of whatever kickbacks the exchange offered without actually providing the liquidity that the kickback was presumably meant to entice. The third, and probably by far the most widespread, they called “slow market arbitrage.” This occurred when a high-frequency trader was able to see the price of a stock change on one exchange, and pick off orders sitting on other exchanges, before the exchanges were able to react. Say, for instance, the market for P& G shares is 80– 80.01, and buyers and sellers sit on both sides on all of the exchanges. A big seller comes in on the NYSE and knocks

the price down to 79.98– 79.99. High-frequency traders buy on NYSE at $ 79.99 and sell on all the other exchanges at $ 80, before the market officially changes. This happened all day, every day, and generated more billions of dollars a year than the other strategies combined. All three predatory strategies depended on speed, and to speed the Puzzle Masters turned their attention, once they were done with the order types. They were trying to create a safe place, where every dollar stood the same chance. How to do that, when a handful of people in the market would always be faster than everyone else? They couldn't very well prohibit high-frequency traders from trading on the exchange— an exchange needed to offer fair access to all broker-dealers. And, anyway, it wasn't high-frequency trading in itself that was pernicious; it was its predations. It wasn't necessary to eliminate high-frequency traders; all that was needed was to eliminate the unfair advantages they had, gained by speed and complexity. Rob Park put it best: "Let's say you know something before everyone else. You are in a privileged state. Eliminating the position of privilege is impossible— some people always will get the information first. Some people will always get it last. You can't stop it. What you can control is how many moves they can make to monetize it."