ALTERNATIVE DATA ACCUMULATION, INVESTMENT MANAGEMENT AND THE EVER-PRESENT SPECTRE OF INSIDER TRADING LIABILITY– SHOULD HEDGE FUNDS BE CONCERNED ABOUT TRADING ON SCRAPED DATA?

Florian N. Kamp*

ABSTRACT

Technological advances have made it possible to scour vast arrays of data in the digital world with algorithms. Investors, in particular hedge funds, are spearheading this technology as means for investment research. In the discussion of this growing trend, the spectre of potential insider trading always looms large and is oft-cited, but seldom analysed in detail. This article looks closely, while trying to be mindful of real-world practices, at the state of play in insider trading doctrine with regard to investments made in reliance on scraped data. Additionally, the article clearly lays out the arguments for and against regulating data scraping via insider trading law – bringing to the forefront the policy concerns which may well be underlying future regulatory and judicial activity in this area –, focusing on incentive mechanisms.

As for the outcome regarding current legal doctrine, utilizing scraped data for investment research will only rarely result in insider trading liability. On the policy side, the arguments against policing any and all data protection violations with insider trading doctrine win out. Bringing the heavy hammer of insider trading down on investors relying on scraped data is ill-suited for likely policy goals, would disincentivize progressive thinkers as well as fossilize market dominance of data giants, and impair the free market equilibrium the U.S. economy is built on.

* PhD, LL.B., Bucerius Law School; LL.M., Harvard Law School. The author would like to thank Holger Spamann and Manish K. Mittal, who piqued the author's interest in this topic with their absorbing style of teaching the course "Hedge and Private Equity Funds: Law and Policy" at Harvard Law School in the Spring of 2019. The idea for this paper was born out of a deliverable for that course.

I. INTRODUCTION

A. Exposé

Drawing on almost every aspect of life, companies, governments and individuals gather a constant stream of data that is published in some publicly available form on the World Wide Web. Much of that data (experts estimate that the overwhelming majority of data that exists worldwide has been created in the last two years[1]) lies about – and takes up space – like trash. Much like trash, however, the ever-inventive business world has found a way to monetize this apparent wasteland of only seemingly useless information. Data vendors are now scraping the internet for all kinds of information, packaging that data in a way that it can be easily transferred and selling that data off to whomever is interested. As with some other trends[2], hedge funds are spearheading this innovative new area, as they – through their fee structure – are the ones most incentivized to seek out new ways to maximize profit.[3]

Macroeconomic trading in currencies and government bonds that relies, among other things, on national statistics can be helped by data that can serve

---

1. *See* Avi Salzman, *Your Personal Data Is Being Used by Investors. Here's the Potential — and the Risks*, BARRON'S (November 30, 2018), https://www.barrons.com/articles/how-big-investors-use-your-personal-data-to-play-the-stock-market-1543627499 [https://perma.cc/4R8A-2X2J] (quoting Tobias Moskowitz, a professor at Yale School of Management: "[W]e produced more data last year than we did in the whole history of humanity").

2. *See, e.g.,* Kendall R. Pauley, *Why Salman is a Game-Changer for the Political Intelligence Industry*, 67 A. U. L. REV. 603, 635–36 (2017) (detailing the use of political intelligence by hedge funds).

3. *See* Alan Crane, Kevin Crotty & Tarik Umar, *Hedge Funds and Public Information Acquisition* 3 (April 23, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3127825 [https://perma.cc/PQ6F-P6UR] ("[S]crapers [hedge funds that systematically use computer programs to gather large quantities of public filings automatically from the SEC website] earn 1.8% higher annualized abnormal returns than non-scrapers"); DELOITTE CTR. FOR FIN. SERVS., ALTERNATIVE DATA FOR INVESTMENT DECISIONS: TODAY'S INNOVATION COULD BE TOMORROW'S REQUIREMENT 3 (2017), https://www2.deloitte.com/content/dam /Deloitte/us/Documents/financial-services/us-fsi-dcfs-alternative-data-for-investment-decisions.pdf [https://perma.cc/GUV9-4A2D] ("Hedge funds have been in the foreground of alternative data innovation"); Matt Egan, *How hedge funds use drones, satellite images and web scraping to gain an edge*, CNN BUSINESS (July 10, 2019), https://edition.cnn.com/2019 /07/10/investing/hedge-fund-drones-alternative-data/index.html [https://perma.cc/VD3X-FQ26] ("The biggest bucket of alternative data is scraping publicly available data from the web."); Jen Wieczner, *How Investors are Using Social Media to Make Money*, FORTUNE (December 7, 2015), https://fortune.com/2015/12/07/dataminr-hedge-funds-twitter-data/ [https://perma.cc/JWP5-FYGZ] (quoting Matthew Granade, chief data analyst at Point72, a hedge fund: "Overall, I think this is a golden age for new investment data sources.").

as a stand-in for inflation rates (e.g. online retailers' general price trends) or by information shining a light on the development of trade balances (port statistics) and trading in shares of individual companies (or sector-specific trading) can be improved by data signaling product interest and, more specifically, sales performance (e.g. price setting at major online retailers; satellite imagery depicting departure of delivery trucks or arrival of customers; foot traffic trends surrounding major retail locations deduced from location data of smartphone users).[4] According to some reports, investment managers (including hedge funds) accounted for around five percent of all internet traffic in 2018 in their search for data underpinning their trading strategy and hedge funds are expected to spend about two billion dollars – and growing – on gathering or purchasing such data.[5] That is to be seen against the backdrop that companies are becoming increasingly cautious about how detailed they want their investor disclosure data to be as they fear short-termist activists reacting to the tiniest aberration ever more.[6] Given the size of this market and the significance of the method as the "next big trend" in trading, the legality of trading on scraped data is of paramount importance to hedge funds – being frequent targets of SEC and DOJ

4. *Wary scouts, Hedge funds worry about the legal risks of using "alternative" data*, THE ECONOMIST (June 21, 2018) [hereinafter THE ECONOMIST], https://www.economist.com/finance-and-economics/2018/06/21/hedge-funds-worry-about-the-legal-risks-of-using-alternative-data [https://perma.cc/RL56-K2GL]. For the slogan of an emerging data analytics firm, see Battlefin, https://www.battlefin.com [https://perma.cc/7VRM-JL9C]: "Using Geolocation to understand customer movement and buying habits, Satellite Imagery to track parking lots, applying Sentiment indicators to interpret news quickly and other methods are helping identify investment ideas." According to a survey of around 70 investment managers who, individually, had between $100 million and $1 trillion assets under management, the six items highest on their "Alternative Data Wishlist" were "Logistics Data, Evaluated Prices, Private company data, Supply chain risk data, Historical credit score data [and] Geolocation data." GREENWHICH ASSOCS., *Alternative Data for Alpha* 6 (January 31, 2017), https://www.greenwich.com/equities /alternative-data-alpha [https://perma.cc/AWF8-83SP].

5. Bradley Saacks, *Hedge funds will spend $2 billion on web-scraping software to gain an edge, and it's part of an investing gold rush*, BUSINESS INSIDER (February 11, 2019), https://www.businessinsider.com/web-scraping-by-hedge-funds-is-growing-rapidly-2019-2 [https://perma.cc/QEY3-KX48]; *see also* DELOITTE CTR. FOR FIN. SERVS., *supra* note 3, at 1 ("Alternative data will likely transform active investment management (IM) over the next five years"). In 2013, scraping is reported to have accounted for a quarter of all Internet activity. Myra F. Din, *Breaching and Entering: When Data Scraping Should be a Federal Computer Hacking Crime*, 81 BROOK. L.REV. 405, 440 (2015).

6. *See* Annie Gaus, *Apple to Stop Breaking Out iPhone Unit Sales -- Investors Aren't Thrilled*, THESTREET (November 2, 2018), https://www.thestreet.com/technology/apple-won-t-break-out-iphone-unit-sales-and-investors-aren-t-thrilled-14767050 [https://perma.cc/VCZ6-Q95F] (reporting on Apple's decision to no longer publish individualized sales figures for iPhones, Macs and iPads from December 2018 on (as well as the precipitous stock drop-off following the announcement of the news in an investor relations call)).

enforcement actions[7] – and has far-reaching policy implications.

SEC v. Dot9[8]: DataPortal, a data processing company founded by two ambitious coders with a background in investment analytics, regularly gathers information off the internet with a scraping algorithm, packages the information according to industry needs and subsequently sells those packages to professional investors, among them Dot9, a hedge fund. In one instance, DataPortal accessed customerreview.com, a website dedicated to providing a grass-roots platform for customer complaints regarding commercial merchandise, and scraped the websites' various threads with the aim of capturing bug and malfunction trends of various technology companies' newest releases. The Terms of Use of customerreview.com state the reviews are designated for public access and that any scraping for commercial purposes is prohibited. Each user/reader is also asked to identify itself via CAPTCHA, a mechanism to weed out bot activity which DataPortal's scraping schema is able to circumvent. In addition to that, DataPortal purchased a set of user location data from LocalCookBook, a popular app that generates regional recipes according to the user's location. Users of the app have all accepted the Terms of Use, which state in the fine print, among other things, that LocalCookBook may use aggregated and anonymized location data for commercial purposes. Finally, DataPortal, its analytics services regularly being contracted for by major technology companies, compiles a statistic showing trends in what percentage of recorded Apple Store visitors buy products versus those who merely report faulty merchandise at the Genius Bar using data that DataPortal had access to for a store utility study it was working on for Apple. The portfolio

7. *See, for example,* United States v. Newman, 773 F.3d 438, 442 (2d Cir. 2014) (involving hedge fund managers for Diamondback Capital and Level Global Investors). *See more generally* Jon Eisenberg, *Insider Trading Law After Salman*, HARV. L. SCH. F. ON CORP. GOV. AND FIN. REG. (January 18, 2017), https://corpgov.law.harvard.edu/2017/01/18/ insider-trading-law-after-salman/ [https://perma.cc/V8Q6-AG66] ("Hedge Fund managers have been among the most frequent targets in both criminal and civil insider trading cases. [. . .] For the period 2010 to 2014 alone, the SEC's "spotlight" on insider trading includes cases against nearly 40 hedge fund managers, hedge funds, and those who allegedly tipped them.").

8. This is a hypothetical case example purely designed to better illustrate some of the issues that are being analysed below. Any factual similarities to actual individuals, corporations or behavioral patterns thereof are unintended.

> manager (PM) at Dot9 purchased all three sets of data with an exclusivity assurance by DataPortal, with the purchase agreement containing an assurance by DataPortal that all information subject to the agreement has been obtained lawfully, being fully aware of where the data came from generally but without specific knowledge of DataPortal's scraping process (even though the level of detail that DataPortal was able to provide on the buying/complaint-lodging customer split did strike him as "odd" in an email to one of his associates, especially given the fact that DataPortal had shared with him the good news of a "fat new contract with Apple"), and uses them, together with his own analysis and other publicly available information to "short" Apple stock for the upcoming quarter as the data purchased from DataPortal shows a sharp increase in customer complaints for a recently released gadget[9], a statistically significant drop-off in foot traffic around Apple stores following the first widely reported malfunctions and an ever-increasing percentage of Apple Store visitors that report product faults. As expected by the PM, Apple stock dips significantly when the new quarterly report (Form 10-Q) is publicly filed with the SEC, and Dot9 is able to generate a fifty million dollar profit on its initial investment of two million dollars through shortselling in a timespan of three months. The SEC is opening an investigation into the matter as it suspects Dot9 (through its PM) has committed insider trading through the use of material, non-public information in executing its securities trades.

The compatibility of data scraping with current insider trading laws is a topic that has garnered significant attention in the wake of the evolution of big data and enhanced government scrutiny in this area: Jonathan R. Streeter, a partner at Dechert LLP, was recently interviewed by Newsweek saying:

> There are some of these data sets that start to look like, wow, someone is really going to have a huge advantage if they have this data set. If you're a hedge fund you may be able to buy this data set and ordinary investors don't have

9. For a real-world example, see Wieczner, *supra* note 3. "Irish research firm Eagle Alpha, for example, digested 7,416 comments on a Reddit gaming thread in October to predict that Electronic Arts (EA) would sell more of its new Star Wars videogame than it had projected; Electronic Arts soon raised its sales forecast, citing "excitement" over the game." *Id.*

> access to that information. And that is material, non-public information about product sales that a public company is going to announce at its next quarterly earnings.[10]

This article will take a look at both the current state of play in insider trading law as well as at the possible policy arguments for and against allowing scraping practices from a doctrinal perspective.

B. Structural Approach

The approach to structuring this topic will first focus on proper definitions of the key terminology, i.e. "insider trading" and "scraped information." The rest of the paper will then be divided into two parts: The first half will focus on whether current practices of data scraping (some hypotheticals will be provided and discussed) can actually lead to insider trading liability. Three areas that will be focused on will be the question of non-public information (since data scraping revolves essentially around a scouring of the public domain[11]), the standard of materiality (since it is unclear if a "reasonable investor would view it as significantly altering the 'total mix' of information available"[12] when only data accumulated by "big data" is concerned[13]) and the issue of whether scraped data can be misappropriated under the common definition (this will depend on whether courts would be willing to equate the terms of use of a website – which oftentimes include safeguards against scraping – to the creation of a fiduciary

10. Ian Allison, Big Data, *Big Problem: Could Wall Street See Insider Trading Lawsuits Over Selling Data Sets?*, NEWSWEEK (November 10, 2017), https://www.newsweek.com/could-wall-street-see-first-legal-action-selling-data-sets-682188 [https://perma.cc/Z8E7-H2XP]. *See also* Egan, *supra* note 3 (quoting Justin Zhen, co-founder of Thinknum, an alternative data provider: "You can wait until companies announce earnings and the whole world will know how companies did. Or, you can know two months in advance."); Kris Kappel & Liam Reilly, *Consider Potential Risks Of Scraping Publicly Available Data*, LAW360 (November 13, 2018), https://www.law360.com/articles/1093344/consider-potential-risks-of-scraping-publicly-available-data [https://perma.cc/BLN4-6VLX] ("So, as the internet continues to evolve, dislocating one industry after another, there is a remarkable irony at the heart of it all — the legal ambiguity of data scraping.).

11. *See*Shaw Horton, *A Fund Manager's Roadmap to Big Data: Its Acquisition and Proper Use (Part Two of Three)*, 3 THE HEDGE FUND LAW REPORT 6 (2018), https://www.lowenstein.com/media/4297/hflr_a-fund-manager-s-roadmap-to-big-data_its-acquisition-and-proper-use.pdf [https://perma.cc/9SKK-JX44] ("[A]ll that managers would have left is that the information obtained is public").

12. Basic Inc. v. Levinson, 485 U.S. 224, 231–32 (1988).

13. Peter Altman, Kelly Handschumacher & Jennifer Hustwitt, *Big Data and the Risks of Insider Trading*, 50 SEC. REG. L. REP. 426, 426–27 (March 19, 2018), https://www.akingump.com/images/content/6/5/v2/65585/spBigData-SRLR-March-19-2.pdf [https://perma.cc/D58K-84RG].

relationship between user and usee[14]).

The other half will be devoted to the policy question of whether data scraping might lead to insider trading liability in the future through shifts in enforcement brought about by a policy impetus. One could question whether trading profits should be acquired on the back of private personal data. Similarly, one can doubt whether the perception of what "non-public" is, needs to be changed in the face of such an overwhelming amount of technically freely available data that only entities with exorbitant computing capabilities can actually make sense of it. On the other hand, there is something to be said for letting companies leverage their data computing capabilities as anything else would constrain their ability to make money. Additionally, threatening SEC enforcement against traders who employ professional information-obtaining strategies may seem dubious as that is basically their job description. Furthermore, one would need to look if such insider trading liability would lead to significantly impeding trading overall (likely a net negative for society as a whole) as the boundaries between legally obtained and scraped information become too blurry for traders to be confident that they know where the line" is – and the efficacy and foreseeability of the legal regime governing securities trades may be put at risk for questionable benefit.

II. TERMINOLOGY

A. Insider Trading

Insider trading liability can arise under the antifraud provisions of Section 10(b) and Rule 10b-5 of the Exchange Act.[15] According to Section 10(b) of the Securities Exchange Act of 1934, as amended, it is unlawful

> [t]o use or employ, in connection with the purchase or sale of any security registered on a national securities exchange or any security not so registered, or any securities-based swap agreement1 any manipulative or deceptive device or contrivance in contravention of such rules and regulations as the Commission may prescribe as necessary or appropriate in the public interest or for the

14. Horton, *supra* note 11, at 6 (claiming that courts regarding the breach of a website's terms of use through the use of deception to be sufficient for insider trading liability to "not be an enormous leap").

15. For a primer, see *Insider Trading*, SEC, https://www.sec.gov/fast-answers/answersinsiderhtm.html [https://perma.cc/Y63P-KMUN].

protection of investors.[16]

According to Rule 10b-5, which the SEC promulgated to colour Section 10(b), it is illegal for any person, "[t]o employ any device, scheme or artifice to defraud," or "[t]o engage in any act, practice, or course of business which operates or would operate as fraud or deceit upon any person [. . .] in connection with the purchase or sale of any security."[17] As the SEC's authority to promulgate a regulation rests on the will of Congress as incorporated in the statute, Rule 10b-5 cannot go beyond what Section 10(b) intended.[18]

The general requirements for insider trading developed under that highly abstract regulatory regime[19] are that someone (1) misappropriated information from someone else to whom he[20] owed a fiduciary duty of sorts, (2) the trader possessed material, non-public information (3) in connection with the sale or purchase of securities and (4) acted with scienter towards requirements (1)-(3). The laws against insider trading are to be interpreted as "enacted for the purpose of avoiding frauds [on the market], not technically and restrictively, but flexibly to effectuate its remedial purposes."[21]

B. Data Scraping

The Oxford Dictionary defines data scraping as: "Extracting large amounts of data from an online source (often using an automated tool), especially where it is then reproduced somewhere else. Search engines routinely do this in ways which benefit web publishers, but in some cases it is a malicious practice."[22] The literal definition of the word in an acclaimed dictionary partially connotes a somewhat underhanded nature of the practice in question.

On a more technical level, Wikipedia defines the practice as involving a "technique in which a computer program extracts data from human-

16. 15 U.S.C.A. § 78j (West).

17. 17 C.F.R. § 240.10b–5.

18. Ernst & Ernst v. Hochfelder, 425 U.S. 185, 212–14 (1976).

19. *But see* Miriam H. Baer, *Insider Trading's Legality Problem*, 127 YALE L. J. FORUM 129, 137 (June 19, 2017) ("[I]nsider trading clearly registers at the end of the spectrum where legislative definition is murky at best").

20. In the following, the use of "he" is also meant to refer to the pronouns "she" and "they".

21. SEC v. Capital Gains Research Bureau, Inc., 375 U.S. 180, 195 (1963).

22. *Scraping,* DANIEL CHANDLER & ROD MUNDAY, A DICTIONARY OF SOCIAL MEDIA (Oxford Univ. Press 2016).

readable output coming from another program."[23]

Speaking more generally and drawing on the definitions mentioned above, data scraping concerns the process of obtaining information through scouring the internet or other open data channels (i.e. the publicly available space) with search or order algorithms[24] – rather than manually perusing them[25] – that enable the user to locate information that can be of economic (trading) value to him (e.g. sales data of specific companies/industry areas, search preferences of customers, mention of specific terms in forums etc.).[26] Publicly searchable (or exploitable) data may nowadays include information that one intuitively considers private, for example, the transaction feed and live data flow of certain apps.[27] So-called data brokers regularly access information made publicly available by state and federal governments, scrape social media and commercial sites, purchase proprietary data sets and then sell that accumulated information to interested parties, among them hedge funds.[28]

---

23. *Data scraping*, WIKIPEDIA, https://en.wikipedia.org/wiki/Data_scraping [https://perma.cc/KH5Z-G8BG] (last visited Feb. 7, 2020).

24. For a practical guide, compare Rasha Ashraf, *Scraping EDGAR* with *Python*, 92 J. EDUC. FOR BUS. 179 (2017), and Michael T. Braun, Goran Kuljanin & Richard P. DeShon, *Special Considerations for the Acquisition and Wrangling of Big Data*, 21 ORG. RES. METHODS 633, 639–40 (2018).

25. *See* Sandvig v. Sessions, 315 F. Supp. 3d 1, 16 (D.C. 2018) ("Scraping is merely a technological advance that makes information collection easier; it is not meaningfully different from using a tape recorder instead of taking written notes, or using the panorama function on a smartphone instead of taking a series of photos from different positions.").

26. See *Ryanair DAC v. Expedia Inc.,* No. C17-1789RSL, 2018 WL 3727599, at *1 (W.D. Wash. Aug. 6, 2018) for a succinct practical example. "The basis for this suit is the way Expedia gets Ryanair's flight and price information. Ryanair alleges that Expedia employs a program—known, among other names, as a "screen scraper"—to automatically gather (or "scrape") data from the Ryanair website. [. . .] The scraper mimics a customer to access the website, sifts through its code, and extracts relevant information about flights, seats, and prices." *Id.* For more detail, see Din, *supra* note 5, at 410–13; Jeffrey K. Hirschey, *Symbiotic Relationships: Pragmatic Acceptance of Data Scraping*, 29 BERKELEY TECH. L.J., ANN. REV. L AND TECH. 897, 903–06 (2014).

27. Recently, a tech-savvy Venmo user was able to scrape an immense volume of transaction details from its data feed. Dan Salmon, *I Scraped Millions of Venmo Payments: Your Data Is at Risk*, WIRED (June 26, 2019), https://www.wired.com/story/i-scraped-millions-of-venmo-payments-your-data-is-at-risk [https://perma.cc/2UA6-73XK].

28. FED. TRADE COMM'N, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY 46–47 (May 2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf [https://perma.cc/Y6G2-LSDW]; Jennifer Valentino-DeVries, Natasha Singer, Michael H. Keller & Aaron Krolik, *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*, N.Y. TIMES (Dec. 10, 2018), https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html [https://perma.cc/MTG7-VQNZ]. In 2018, the market value of consumer data trade was valued at $19 billion in

III. LEGAL ANALYSIS – ARE TRADES BASED ON SCRAPED DATA AT AN ACUTE RISK OF BEING CONSIDERED INSIDER TRADING VIOLATIONS?

The legal analysis will follow the structure of insider trading prerequisites. More specifically, it will focus on the materiality of scraped data, whether such information can be considered non-public and if, and to what extent, data scraping can induce a breach of fiduciary duties. Finally, the paper will jointly look at the personal benefit requirement of insider trading, in case the trader and the information gathering party are not the same, details of the tippee's liability and the extent to which scienter plays into both.

A. Materiality of the Information

1. Basic Rationale

One of the main elements of civil liability under § 10(b) and Rule 10b-5 is that the information in question must have been material.[29] The judicial standard for materiality of information "there must be a substantial likelihood that the [. . .] fact would have been viewed by the reasonable investor as having significantly altered the 'total mix' of information."[30]

The courts have previously found that a company's revenue information can be material under this standard as a "reasonable inference [can be drawn] that a reasonable investor would see the obvious connection between increased revenues and the likelihood of increased profits."[31] If the information is, however, only regarding such a small portion of revenue that it cannot possibly have a meaningful impact on the overall outlook of the company, it will not be considered material.[32] While the SEC has

---

the U.S. alone. Louise Matsakis, *The WIRED Guide to Your Personal Data (and Who Is Using It)*, WIRED (Feb. 15, 2019), https://www.wired.com/story/wired-guide-personal-data-collection/ [https://perma.cc/VP48-B7F3].

29. *See* 17 C.F.R. § 240.10b-5 ("It shall be unlawful for any person [. . .][t]o make any untrue statement of a material fact or to omit to state a material fact necessary in order to make the statements made, in the light of the circumstances under which they were made, not misleading").

30. Basic Inc. v. Levinson, 485 U.S. 224, 231–32 (1988) (quoting TSC Indus., Inc. v. Northway, Inc., 426 U.S. 438, 449 (1976)).

31. Rothberg v. Rosenbloom, 771 F.2d 818, 821 (3d Cir. 1985).

32. *In re* Westinghouse Sec. Lit., 90 F.3d 696, 714–15 (3d Cir. 1996) (denying materiality for information on an item on the balance sheet that amounted to only about 0.5% of the company's overall revenue).

commented more broadly that "filling in the 'interstices in analysis'" should not give rise to insider trading,[33] the U.S. Supreme Court has found that suggested distinction "inherently imprecise" and argued that the line should be drawn as clearly as possible.[34]

2. Application in the Digital World

At first sight, scraped data seems too meaningless on its own to be material: Given that these datasets are often provided as part of a whole bundle of information – a large majority of which will be of no significant use whatsoever, let alone prove material –, their singular expressiveness seems somewhat muted.[35]

That first approximation notwithstanding, even information that is only directly giving insight into a small portion of a corporation's revenue can be material when that data set has implications going beyond the revenue immediately impacted and can be used in conjunction with other (publicly available) data to reach conclusions that would otherwise not be possible. That holds true especially since technological advances have made it possible to piece together information into a cohesive message that would previously have remained meaningless.[36] In the case of Huang, the U.S. Court of Appeals for the Third Circuit opined that credit card data that accounted for around 2.4% of overall sales of the retail companies the analyst then traded on was still likely to be considered material as the analyst in question used that "data (in tandem with publicly available information) to predict total revenue information with greater accuracy than analysts using only publicly available information."[37] In a similar vein, the U.S. Supreme Court held in the context of large-scale data accumulation, that a lack of

33. Dirks v. S.E.C., 463 U.S. 646, 659 n.17 (1983) (quote from the SEC briefs).

34. *Id.*

35. *See* Richard A. Epstein, *Returning to Common-Law Principles of Insider Trading After United States v. Newman*, 125 YALE L. J. 1482, 1523 (2016) ("The value of information varies inversely with the number of people who share it, and the rapidity with which that information is factors into overall markets valuations. [. . .] The abundance of other information further dilute[s] the significance of the information"). With regard to insider trading risks for hedge funds trading on political intelligence, *see* Pauley, *supra* note 2, at 645 ("[B]undling makes it difficult for prosecutors to establish that any one piece of information is material").

36. Nicolas H.R. Dumont, *Sentiment Analysis & Natural Language: Processing Techniques for Capital Markets & Disclosure*, 25 CORP. GOVERNANCE ADVISOR 16, 18 (November/December 2017) (commenting on the materiality threshold in the digital age: "relating what is "material" may be more challenging because counterparties are listening in a way that most humans never intended").

37. SEC v. Bonan Huang, 684 F. App'x 167, 172 n.6 (3d Cir. 2017).

statistical significance does not preclude the information from being material.[38] This means that courts are unlikely to look too kind on a defendant arguing that trading on data otherwise falling under the purview of insider trading laws was only reflective of a holistic analysis where the data in question only played a minor part in filling out "interstices."

Especially in cases where outsize returns are attained by the investors, the courts sometimes also employ basic rationality logic to strengthen the case for the materiality of the information[39]: In Rothberg, the court arrived at the materiality finding by basically extrapolating from the point that "experienced investors" (with hindsight bias, successfully, one needs to add) acted upon the information.[40] In Huang, the court also conflated the enormous returns garnered by the investors with the importance of the information relied upon for those returns.[41]

> SEC v. Dot9: All three data sets purchased by Dot9 are likely to be considered material by the courts as the uptick in customer complaints, the location data focusing on Apple stores and the statistic regarding the split between customers who buy something in Apple stores and those who merely report faults all give the trader a meaningful informational advantage over the average investor. Only with regard to the location data could one argue that the mere trend in foot traffic is not by itself notable. A decrease in foot traffic could also mean less customers lodging complaints in person and/or more online

38. *See* Matrixx Initiatives, Inc. v. Siracusano, 563 U.S. 27, 43 (2011) ("Given that medical professionals and regulators act on the basis of evidence of causation that is not statistically significant, it stands to reason that in certain cases reasonable investors would as well.").

39. *See also* INTEGRITY RESEARCH ASSOCS., MITIGATING LEGAL RISKS ASSOCIATED WITH ALTERNATIVE DATA 9 (January 2018), https://www.integrity-research.com/wp-content/uploads/2018/01/Mitigating-Legal-Risks-Alternative-Data-January-2018-2.pdf [https://perma.cc/V69S-NV7T] ("Alpha-generating information is by definition material. Insider trading prosecutions are all ex-post, allowing prosecutors to focus only on successful trades, and the more successful the trade the more attractive it is to a prosecutor.").

40. *See* Rothberg v. Rosenbloom, 771 F.2d 818, 821 (3d Cir. 1985) ("The best proof of the materiality of [. . .] information [regarding sales orders] is that [. . .] experienced investors found it to be sufficiently material to form the joint venture and to purchase stock when it was depressed in price").

41. *See* SEC v. Bonan Huang, 684 F. App'x 167, 173 (3d Cir. 2017) ("As reflected by Huang's own investment decisions, which [. . .] resulted in a 12,929% three-year return on his investment, the nonpublic Capital One data "significantly altered the 'total mix' of information" in the eyes of a reasonable investor.").

> purchases.[42] Only together with additional information (the internal Apple store statistics) does the data unequivocally support the "shorting" of Apple stock. Given the fact that foot traffic trends can likely be at least a statistically significant proxy for sales trends and therefore would likely seem material to a reasonable investor[43], the more than twentyfold return on investment, and the courts' unwillingness to entertain the notion that information that factually lead someone to make a significant investment which yielded outsize returns was somehow immaterial, such an argument seems unlikely to prevail at trial.

B. Non-Public Nature of the Information

1. Basic Rationale

As a baseline that seems obvious, but is worth repeating here, considering the question at hand revolves around the possible illegality of trading on data essentially gathered off the Internet, it is "axiomatic that trading on public information does not violate Section 10(b)."[44] Information can become public if it is released in a manner that is "designed to achieve a broad dissemination to the investing public generally and without favoring any special person or group."[45] Ordinarily, that means information is public "if it is available to the public through SEC filings, the media, or other sources."[46] Even if only a small number of people know of the information, the information is to be considered public if "their trading has caused the information to be fully impounded into the price of the particular stock" as this renders any market abuse impossible since there are no more

---

42. *See* Brandon Kochkodin, *Parking Lots Don't Tell the Whole Story: The Trouble With Alternative Data*, BLOOMBERG (November 29, 2018), https://www.bloomberg. com/news/articles/2018-11-29/the-trouble-with-using-alternative-data-to-gain-an-investing-edge [https://perma.cc/6DDF-KGAC] (quoting John Chisholm, Co-CEO at Acadian Asset Management: "How consistently does foot traffic translate into retail sales? And even if it does, is the earnings impact already discounted by analysts?).

43. *See* Basic Inc. v. Levinson, 485 U.S. 224, 240 (1988) ("[M]ateriality depends on the significance the reasonable investor would place on the withheld or misrepresented information").

44. United States v. Libera, 989 F.2d 596, 601 (2d Cir. 1993).

45. Dirks v. S.E.C., 463 U.S. 646, 653–54 n.12 (1983).

46. United States v. Contorinis, 692 F.3d 136, 143 (2d Cir. 2012).

(illegitimate) profits to be reaped.[47]

2. Non-Public Nature of Aggregated Public Information?

In relation to trading done by hedge funds, it is fair to assume that the information will most likely not have been "impounded into the price of the particular stock"[48] as there would – from the perspective of the resourceful, evaluating and value-maximizing individual[49] – be no profit (i.e. alpha) in trading on the information in that case.[50] One could, however, argue that a lot of the information is broadly disseminated to the public as scraping usually involves the search of public domains.[51] Even though the specific dataset gathered and arranged may not be available to the public – as the value-add of data analytics lies specifically in arranging the information in such a way that meaningful conclusions can be drawn from it that weren't discernible before[52] – the source of the information, i.e. the World Wide Web in many instances, was and is publicly available. The question then becomes whether the information – as information overload is a common feature nowadays and even considered a potential threat to disclosure obligations[53]

47. United States v. Libera, 989 F.2d 596, 601 (2d Cir. 1993).

48. *Id.*

49. For more details on the REM model, see Michael C. Jensen & William H. Meckling, *The Nature of Man*, 7 J. APPLIED CORP. FIN. 4, 7–10 (Summer 1994).

50. Another area in which the non-public nature of information is debated, is with regards to data gathered through Freedom of Information Act (FOIA) requests, as the availability of such information depends on both a prior request and the payment of a fee. Caroline Banton, *5 Ways to Make Money That Should Be Illegal*, THE PHILADELPHIA INQUIRER ,Apr. 13, 2016, https://www.inquirer.com/philly/business/5_Ways_to_Make_Money_That_Should_Be_Illegal.html [https://perma.cc/392X-KEJP] (quoting Max Galka, a derivatives trader mapping FOIA requests: "I use the term 'not publicly known' rather than 'nonpublic' because technically the information is public, and trading on it is legal. However, since the information is only obtainable using the Freedom of Information Act, for all intents and purposes, it is nonpublic information."). Hedge Funds account for a nontrivial percentage of FOIA request. Max Galka, *Who Uses FOIA? – An Analysis of 229,000 Requests to 85 Government Agencies*, FOIA MAPPER, Mar. 13, 2017, https://foiamapper.com/who-uses-foia/ [https://perma.cc/6QX6-DNG5].

51. Egan, *supra* note 3 ("[S]trategy as it's intended does not involve using confidential information.").

52. *See* Crane, Crotty & Umar, *supra* note 3, at 19 (commenting on the outcome of an abnormal-returns study on hedge fund: "Overall, these results are consistent with hedge fund possessing private information about upcoming events for firms and researching the firms ahead of the public revelation of this information."); Egan, *supra* note 3 (quoting Justin Zhen, co-founder of Thinknum, an alternative data provider: "It's a way for investors to almost spy on management").

53. *See* ELSIE HENDERSON, USERS' PERCEPTIONS OF FINANCIAL STATEMENT NOTE DISCLOSURE AND THE THEORY OF INFORMATION OVERLOAD 115 (ProQuest Dissertations

– can be thought of as public only because its underlying data could be accessed piecemeal by the general public.

(a) First Approximation

There is no case law on the specific question of whether scraped data is considered non-public under insider trading law and the ever-changing nature of technology makes doubling down on an abstract answer difficult.[54] Therefore, the most convincing argument is likely to prevail at trial. While the emphasis on the public origin of the scraped data has some superficial appeal, it seems more sensible to qualify the aggregated data set a new piece of information that was not previously available. In today's ubiquity of data, purposefully gathering data and distilling it to a set that actually conveys a message to the surveyor that the amorphous array of material were not (yet) able to articulate, is a skill so material that it warrants classifying the aggregation as a novel piece of information.[55]

(b) Comparison with EU Legislation

This line of reasoning, which is a product of both technological progress in the field of data analysis and the growing deluge of publicly accessible

---

Publishing 2016) ("Transparency must always be an overarching component of disclosures. As standard setters work to improve financial statement note disclosures more emphasis must be placed on providing succinct disclosures and reducing disclosures adding little value to users through a clearer expectation of the application of materiality. Users need to understand more disclosure does not mean better disclosure or greater transparency, in fact more could cause users to lose sight [of] significant information."); Ricarda Moll, Stefanie Pischl & Rainer Bromme, *Whoever will read it – The overload heuristic in collective privacy expectations*, 75 COMPUTERS HUM. BEHAV. 484, 485 (2017) ("When people are confronted with more information than they can actually process, information tends to be perceived as noise, namely redundant or meaningless information that interferes with the goals or expected signals of the receiving person").

54. *See* Wieczner, *supra* note 3 (quoting Daniel Hawke, former chief of the SEC's market abuse unit: "In a world where information travels very, very fast and through different media, figuring out whether information is public or not is challenging. . . .").

55. *See* Enrico Colombatto & Valerio Tavormina, *Regulating information flows: Is it just? Insider Trading and mandatory-disclosure rules from a free-market perspective*, 46 EUR. J. L. ECON. 205, 214 (2018) (referencing a "Lockean claim in favour of exclusive property rights" for data analysis); Crane, Crotty & Umar, *supra* note 3, at 2 n.3 ("Note that private information is not necessarily illegal insider information but could instead stem from hard-to-get or costly data sources (e.g., satellite data, mobile phone data). . . ."). *See also* DELOITTE CTR. FOR FIN. SERV., *supra* note 3, at 7 ("[A]n important lesson for IM firms to consider is that regulators are taking note of alternative data, and the common definitions of public and private information are in transition. . . .").

information, can arguably be backed up by a comparison with the evolution of EU regulation on the topic: Whereas the original EU Directive on insider trading explicitly, and without any qualification or caveats, exempted trades based on "research and estimates developed from publicly available sources"[56], the revamped Regulation on this topic now cautions traders that thusly derived information is "not per se" inside information leading to insider trading liability and advises investors, rather vaguely, one might add, to "consider the extent to which the information is non-public and the possible effect on financial instruments traded in advance of its publication or distribution, to establish whether they would be trading on the basis of inside information."[57] This explicit backtracking on the question of whether information derived from publicly available data can be inside (i.e. nonpublic) information on the part of the EU legislature shows that this is less of an open-and-shut case than previously thought.

(c) Role of Exclusivity Agreements

As reported by the Financial Times, the Chief Market Intelligence Office of Point72, a hedge fund based in New York, said in response to a question at a student panel as to how Big Data was helping hedge funds get an alpha-edge when everyone has access to the same information: "The great thing about this area is you can arrange deals where you are the only ones who get it."[58] More specifically, he was referring to exclusivity agreements with the data vendors. Since a lot of data distribution companies are pitching

56. Directive 2003/6/EC of the European Parliament and of the Council of 28 January 2003 on insider dealing and market manipulation (market abuse), 2003 O.J. (L 96) 1, 18 ("Research and estimates developed from publicly available data should not be regarded as inside information and, therefore, any transaction carried out on the basis of such research or estimates should not be deemed in itself to constitute insider dealing within the meaning of this Directive.").

57. Regulation (EU) No 596/2014 of the European Parliament and of the Council of 16 April 2014 on market abuse (market abuse regulation) and repealing Directive 2003/6/EC of the European Parliament and of the Council and Commission Directives 2003/124/EC, 2003/125/EC and 2004/72/EC, 2014 O.J. (L 173) 1, 7–8 ("Research and estimates based on publicly available data, should not per se be regarded as inside information and the mere fact that a transaction is carried out on the basis of research or estimates should not therefore be deemed to constitute use of inside information. . . . Market actors must therefore consider the extent to which the information is non-public and the possible effect on financial instruments traded in advance of its publication or distribution, to establish whether they would be trading on the basis of inside information.")

58. Lindsay Fortado, Robin Wigglesworth & Kara Scannell, *Hedge funds see a gold rush in data mining*, FINANCIAL TIMES (Aug. 28, 2017), https://www.ft.com/content/d86a d460-88 02-11e7-bf50-e1c239b45787 [https://perma.cc/3HAD-ZH7Y].

to hedge funds to purchase large data sets, many hedge funds will only agree to contract if there is an exclusivity agreement in place with regard to the specific data sold, as "alternative data" is perceived to be the "next big thing" in generating alpha, which means there is an innate desire on the part of hedge funds to restrict access to such data sets.[59]

> SEC v. Dot9: Through exclusivity agreements with DataPortal, Dot9 protected and perpetuated the private (i.e. non-public) nature of the data. Those exclusivity agreements emphasize the scarcity of the aggregated data sets and therefore their non-public nature.[60]

(d) Prognosis

Better erring on the side of caution in trying to predict the destination of a moving target, investment managers should assume that the SEC might prevail with the argument that scraping and further analysis are capable of creating a new, non-public piece of information. Policy concerns regarding the possible overreach of the insider trading laws appear unwarranted at this stage as the breach of fiduciary duty and the scienter element – additional hurdles for an insider trading case to clear in the U.S. that the EU regulation on insider trading, for example, does not demand in this form[61] – remains as a (meaningful) restriction on the applicability of Section 10(b) and Rule 10b-5.

---

59. *See* CITI BUS. ADVISORY SERVS., BIG DATA & INVESTMENT MANAGEMENT: THE POTENTIAL TO QUANTIFY TRADITIONALLY QUALITATIVE FACTORS 17 (October 7, 2019 12:49 PM), https://www.cmegroup.com/education/files/big-data-investment-management-the-potential-to-quantify-traditionally-qualitative-factors.pdf [https://perma.cc/7Q9S-HYWK] (describing "growing number of instances where a provider is looking to supply data to a sole or limited set of users at significant price premiums"); Crane, Crotty & Umar, *supra* note 3, at 26 ("We investigate whether the positive relation between performance and public information acquisition is due to hedge funds being superior information processors or if it is due them public information in conjunction with private signals about firm values. Several analyses suggest the latter channel is the predominant source of the relation."), 2–3, n.3 ("Note that private information . . . could . . . stem from hard-to-get or costly data sources (e.g., satellite data, mobile phone data). . . .").

60. *See* INTEGRITY RESEARCH ASSOCS., *supra* note 39, at 8 ("Insider trading risks . . . can be heightened by exclusivity agreements.").

61. *See* Merritt B. Fox, Lawrence R. Glosten & Gabriel V. Rauterberg, *Informed Trading and its Regulation*, 43 J. CORP. L. 817, 886–87 (2018) ("[T]he Directive is not hobbled by the personal benefit rule test nor does it require a showing of knowledge by the trader of a prior breach of some duty. . . .").

SEC v. Dot9: The data gathered off of customerreviews.com is publicly available. That opens the door for the argument that – even though its content is undisputedly not "impounded" into the current price of Apple stock as it does not yet reflect the disappointing business outlook revealed in the 10-Q – the underlying information is already accessible by the public. The SEC might argue that the scraping process created a dataset that was so different from the breadth of customer reviews previously accessible, that it gave rise to new information which, in the hands of its creators, became non-public data. Given the ubiquity of information in the digital age, the molding of data into a cohesive novel set likely created a "new" piece of information under Section 10(b) and Rule 10b-5. The private nature of the information was even emphasized by the purchase price and the exclusivity agreement that Dot9 and DataPortal agreed on (why else would one want to pay for and protect exclusive access to a dataset if the information is already considered public?). The data used to disseminate proportions of buying versus complaint-lodging customers originates from inside Apple and is therefore non-public in all aspects. Finally, the location data was initially only gathered by the proprietor of LocalCookBook and was therefore never public.

C. Breach of Fiduciary Duty as a Restriction on Insider Trading Doctrine

1. Basic Rationale

Insider trading liability can rest on either one of two separate doctrines: Under the classical theory of insider-trading liability, the duty breach occurs with respect to the shareholders of the defendant's corporation when a corporate insider or his tippee trade securities of that company on the basis of material, non-public information about said entity.[62] If there is no

62. Salman v. United States, 137 S. Ct. 420, 425 n.2 (2016). See, for example, the two factual components in a recent SEC civil complaint alleging that Apple's in-house chief corporate lawyer traded on information in violation of his duty to shareholders: "1. As a Senior Attorney and Member of Apple's Disclosure Committee, Levoff Was Entrusted with Material Nonpublic Information"; "2. At the Time of His Trading, Apple Had Taken Steps to Prevent

company insider involved, insider trading liability can arise under the so-called misappropriation theory when someone misappropriates information gleaned from a source which the person owes a fiduciary duty to, the illegality of later trading notionally resting on the "fiduciary-turned-trader's deception of those who entrusted him with access to confidential information."[63]

When – as described in the hypothetical of SEC v. Dot9 – the party ultimately executing the securities trade (in this scenario, the tippee) is receiving the information from another person (referred to as the tipper), the duty breach must originally occur between the tipper and either his corporate employer (under the classical theory) or the party that the tipper owes fiduciary responsibilities to (under the misappropriation theory).[64]

> SEC v. Dot9: DataPortal owed a fiduciary duty to its client, Apple, to keep all proprietary information it received from Apple in the course of their business relationship secret and not disclose it to outsiders. By including the statistic regarding the split between Apple store customers who buy and those who just "complain" in their data aggregation sold to Dot9, DataPortal violated said fiduciary duty.

2. Necessity of a Fiduciary Breach in the Context of Scraping?

Having established that a fiduciary breach is generally necessary to justify a finding of insider trading liability, the quandary remains whether that holds true in the context of data scraping as well.

(a) The Case of Dorozhko

Namely, in the case of *Dorozhko*, the Second Circuit found that no precedent existed that prevented it from equating a deceptive hack with an affirmative misrepresentation that made further findings on the existence of

---

Employees, Such as Levoff, from Trading on Material Nonpublic Information." SEC v. Gene Daniel Levoff, 2019 WL 630326 (D.N.J.) (No. 2:19-5536).

63. *See* United States v. O'Hagan, 521 U.S. 642, 652, 656 (1997) ("[M]isappropriator . . . gains his advantageous market position through deception; he deceives the source of the information and simultaneously harms members of the investing public.").

64. *See* Dirks v. SEC, 463 U.S. 646, 664 n.23 (1983) ("Only if there was such a breach [i.e., of the tipper's fiduciary responsibilities towards his employer and the shareholders of said employer] did Dirks, a tippee, acquire a fiduciary duty to disclose or abstain.").

a fiduciary duty unnecessary.[65] The case has generated significant discussion around the necessity of a fiduciary breach requirement for technological data extraction.[66] While the necessary degree of relationship between insider trading and the ever-changing chameleon of fiduciary obligations poses an intriguing question from an academic perspective,[67] the case of *Dorozhko* remains an outlier in U.S. case law on insider trading.[68] Even in the case that was decided, the court left open the question as to whether the hacking at hand constituted affirmatively deceptive behavior that supposedly suffices for the stretched definition of insider trading.[69] Given the factual difficulty that arises when one is to apply the term of deception – a concept that would

65. *See* S.E.C. v. Dorozhko, 574 F.3d 42, 49 (2d Cir. 2009) ("[T]he SEC argues that defendant affirmatively misrepresented himself in order to gain access to material, nonpublic information, which he then used to trade. We are aware of no precedent of the Supreme Court or our Court that forecloses or prohibits the SEC's straightforward theory of fraud. Absent a controlling precedent that "deceptive" has a more limited meaning than its ordinary meaning, we see no reason to complicate the enforcement of Section 10(b) by divining new requirements.").

66. *See* Abraham C. Bloomenstiel, *Proprietary Data Feed and Colocation-Enabled High Frequency Trading: Troubling Paradoxes and Difficult Truths*, 45 SEC. REG. L.J. 147, 157 (2017) (finding that "duty-based approaches to insider trading may be eroding"); Brian A. Karol, *Deception Absent Duty: Computer Hackers & Section 10(B) Liability*, 19 U. MIAMI BUS. L. REV. 185, 196 (2011) ("At least two federal circuit courts are split on the issue of whether the Supreme Court has authoritatively construed Section 10(b) to require a breach of a fiduciary duty for any "device" to be "deceptive".); Elizabeth A. Odian, *SEC v Dorozhko's Affirmative Misrepresentation Theory of Insider Trading: An Improper Means to a Proper End*, 94 MARQ. L. REV. 1313, 1333 (2011) ("new theory of insider [. . .] that substitutes the longstanding fiduciary duty requirement with a test for common law fraud"). *See also* Eric C. Chaffee, *The Supreme Court as Museum Curator: Securities Regulation and the Roberts Court*, 67 CASE W. RES. L. REV. 847, 885–86 (2017) ("[Supreme] Court in Dirks acknowledged that deception is at the core of an insider trading violation, rather than a breach of fiduciary duty").

67. *See* Chaffee, *supra* note 66, at 887 ("[D]eception based theory of insider trading is also more in keeping with the history and intent of Rule 10b-5").

68. *See* Don Butterworth, *SEC v. Dorozhko: Alternative Data, Web Scraping, and 10(b) Fraud*, COLUM. BUS. L. REV. ONLINE (Aug. 13, 2019), https://journals.library.columbia.edu/index.php/CBLR/announcement/view/198 [https://perma.cc/Z47D-7P46] ("In the nearly ten years since Dorozhko, no court has taken up the challenge of determining the extent of potential liability for "outsider traders" who fraudulently misrepresent their identities to obtain MNPI."). For a critical view towards the status quo of (Supreme Court) precedents in this area, see Chaffee, *supra* note 66, at 885: "Considering how much fiduciary duty law has evolved since Dirks and how much it is likely to evolve in the future, one must wonder why the Roberts Court continues to build on this unstable foundation for insider trading regulation."

69. *See* SEC v. Dorozhko, 574 F.3d 42, 51 (2d Cir. 2009) ("Having established that the SEC need not demonstrate a breach of fiduciary duty, we now remand to the District Court to consider, in the first instance, whether the computer hacking in this case involved a fraudulent misrepresentation that was "deceptive" within the ordinary meaning of Section 10(b).").

seem to presuppose, at a minimum, the ability to think (if not free thought) – to a machine-based algorithm,[70] this standard will give little insight into what courts might to with it in the future. So it really all depends on the specificities of the hack as well as the deciding court.[71]

(b) The Limited Impact of Dorozhko

Even if one were to accept the notion that hacking obviates the need for the finding of a fiduciary breach,[72] the question still remains whether that precedent at all applies to data scraping practices. A definitive answer to that question would require a technical analysis of the scraping practice at hand that can and should not be anticipated without a fact pattern in mind. Generally, however, while there might some overlaps between hacking and scraping whenever the scraping process is preceded by a security breach – since, technically, every hack with the intent to glean information from a system is followed by an informational scrape of some sort –, for the most part scraping should be hack-free. The scraping process is predicated on grabbing information off of publicly available sources, rendering any system breach unnecessary. Whether automated scraping is in compliance with a system's terms of use – which may oftentimes be questionable, especially if the availability of that information for the general public is factually limited to highly sophisticated tech professionals who "know where to look" – is not the same issue as whether that process constitutes hacking, the latter clearly implying the forced entry into a closed-circuit system or the intentional circumvention of technological barriers through the technological equivalent of force.[73] In scraping scenarios where it might be applicable and one would rely on *Dorozhko* as the leading case in this area, the scraping process would nevertheless need to be affirmatively deceptive – as this is the only precedent

70. While the literature is torn on whether computers can "think" or not, the surest way to legitimize the term of "deception" in this context would be to extend personhood – already granted to corporations by the U.S. Supreme Court – to computers. *See* Farid Sharaby, *Computer Hacking as a "Deceptive Device": Why the Courts Must Give Computers Legal Consciousness to Hold Hackers Liable for Insider Trading*, 42 MCGEORGE L. REV. 929, 941–53 (2011).

71. *See* Karol, *supra* note 66, at 214 ("Courts have not clarified whether computer hacking is "inherently deceptive", as the SEC alleges, or if only certain types of hacking are deceptive.").

72. *See* Karol, *supra* note 66, at 206 ("may now be possible for non-fiduciaries to violate Section 10(b)'s prohibition").

73. *See* SEC v. Dorozhko, 574 F.3d 42, 44 (2d Cir. 2009) ("[A]n anonymous computer hacker attempted to gain access to the IMS earnings report by hacking into a secure server at Thomson prior to the report's official release").

Dorozhko has actually set with regard to a softening of the "fiduciary breach"-requirement[74] –, meaning it has to satisfy a somewhat vague standard[75] of technological fraud. Thus, the infamous precedent of *Dorozhko* currently has limited impact on the general necessity of a fiduciary breach.[76]

3. Website Scraping as a Fiduciary Breach?

Another question is whether data that is gathered under violation of the Terms of Use of a website fulfils the requirement of information gleaned (and subsequently either used for trading purposes or transmitted to a third party for said purpose) in breach of a fiduciary duty. In this scenario, the data will usually not have originated from a corporate insider, meaning insider trading liability can only arise under the misappropriation theory. Outlawing a trade under that theory rests on the notion, however, that "undisclosed misappropriation of [. . .] information, in violation of a fiduciary duty, [. . .] constitutes fraud akin to embezzlement."[77] Therefore, to hold someone liable under insider trading law for accessing outwardly public information on a website, the use of that website with a scraping algorithm must somehow be qualified as deceptive as that is the underlying rationale of Section 10(b).

(a) Terms of Use as Fiduciary Relationship?

One way to potentially do that would be to equate a website's terms of use – that oftentimes bar the commercial harvesting of its content[78] – with a fiduciary relationship between the user and the operator of a webpage. If, for example, a Venmo user exploits a loophole in the app and its website to

74. *See* Bloomenstiel, *supra* note 66, at 158 (calling Dorozhko an "expansion of the legal theory underlying insider trading").

75. Butterworth, *supra* note 68 ("The Second Circuit…offered only some cryptic guidance"); Karol, *supra* note 66, at 206 ("[The] breadth of the meaning of 'deceptive' under Section 10(b) still remains unclear").

76. Similar conclusion drawn by Bloomenstiel, *supra* note 66, at 158. "Though the 2nd Circuit's holding in Dorozhko weakens breach-of-duty as an essential element for insider trading liability, it does not signal a retreat to an "information parity" approach to insider trading." *Id.*

77. United States v. O'Hagan, 521 U.S. 642, 654 (1997).

78. *E.g.* Clause 8.2.(b) of *User Agreement*, LINKEDIN (last updated Jan. 6, 2020), https://www.linkedin.com/legal/user-agreement?trk=homepage-basic_footer-user-agreement#dos [https://perma.cc/BTJ3-3QF2] ("You agree that you will not: [. . .] Develop, support or use software, devices, scripts, robots, or any other means or processes (including crawlers, browser plugins and add-ons, or any other technology) to scrape the Services or otherwise copy profiles and other data from the Services").

scrape detailed, though anonymous, transaction data for millions of transactions,[79] that behavior will most likely be in violation of some term of use.[80] Even though no insider trading was at stake here, LinkedIn has tried to argue something similar when it (unsuccessfully) tried to shut out a data analytics firm that scraped its public page in breach of its explicit will for the purpose of compiling data on a companies' most "volatile" (i.e. likely to leave) employees.[81] The case of LinkedIn – that was recently decided in hiQ's favor by the Ninth Circuit[82] – has garnered close attention from hedge funds which are worried that one of their prime sources for alpha-generating trading information – public webpages ripe for a scraping algorithm – might be cordoned off in the future.[83]

79. A computer science student has successfully managed to do so. *See* Salmon, *supra* note 27 (documenting an individual scraping transaction data off of Venmo).

80. *See User Agreement*, *Restricted Activities*, VENMO (last updated Jan. 27, 2020), https://venmo.com/legal/us-user-agreement/ [https://perma.cc/5A77-VNAE] ("In connection with your use of our websites, your Venmo account, the Venmo services, or in the course of your interactions with us, other customers, or third parties, you must not [. . .] Use an anonymizing proxy; use any robot, spider, other automatic device, or manual process to monitor or copy our websites without our prior written permission").

81. hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1106 (N.D. Cal. 2017), *aff'd and remanded*, No. 17-16783, 2019 WL 4251889 (9th Cir. Sept. 9, 2019) ("[W]hile the information that hiQ seeks to collect is publicly viewable, the posting of changes to a profile may raise the risk that a current employee may be rated as having a higher risk of flight under Keeper even though the employee chose the Do Not Broadcast setting. hiQ could also make data from users available even after those users have removed it from their profiles or deleted their profiles altogether. LinkedIn argues that both it and its users therefore face substantial harm absent an injunction; if hiQ is able to continue its data collection unabated, LinkedIn members' privacy may be compromised, and the company will suffer a corresponding loss of consumer trust and confidence.").

82. hiQ Labs, Inc. v. LinkedIn Corp., No. 17-16783, 2019 WL 4251889 (9th Cir. Sept. 9, 2019).

83. Bradley Saacks, *Hedge funds are watching key lawsuit involving LinkedIn to see if they can spend billions on web-scraped data*, BUSINESS INSIDER (Mar. 14, 2019), https://www.businessinsider.com/hedge-funds-watching-linkedin-lawsuit-on-web-scraped-data-2019-3 [https://perma.cc/D8UH-FT7A]; Tristan Greene, *We should be getting paid to use Facebook and Google*, TNW (Mar. 9, 2018), https://thenextweb.com/artificial-intelligence/2018/03/09/we-should-be-getting-paid-to-use-facebook-and-google/ [https://perma.cc/DP5H-USPP] ("could have huge implications throughout the technology world"); Tony Hughes, *Moody's Analytics Economist: Why the LinkedIn Data Case Is a Lose-Lose Situation*, FORTUNE (Mar. 16, 2018), https://fortune.com/2018/03/16/linkedin-hiq-labs-data-case/ [https://perma.cc/LDE5-486M] ("[E]specially cruel choice in cases like this. We can make our data freely available and have no one bother to collect it. Or, we can bestow ownership rights on the data and potentially miss out on beneficial insights gleaned from its analysis.").

(b) Terms of Use as Access Restriction?

Considering the ubiquity of potential users, one could question whether using a publicly accessible website in any shape or form that does not involve hacking the website's security's protocol[84] can breach a fiduciary relationship between user and usee[85]: The mere presence of technological barriers against bot-centric scraping such as CAPTCHA-technology ("to verify that you are not a robot") does not automatically remove the content from the public sphere, i.e. such scraping does not, in and of itself, constitute unauthorized access, for example, under the Computer Fraud and Abuse Act (CFAA).[86] Those doubts are reinforced by the fact that even for members of

84. The act of hacking—e.g. giving a computer system, through technological deception, the false impression that one is "authorized" to access data—is in itself deceptive, rendering further inquiries into fiduciary breaches arguably superfluous. *See* SEC v. Dorozhko, 574 F.3d 42, 49–51 (2d Cir. 2009). The Court left open whether "exploiting a weakness in an electronic code to gain unauthorized access is 'deceptive', rather than being mere theft." For further analysis on the Second Circuit's definition of 'deceptive' and whether that meshes with the current Supreme Court precedents, *see* Karol, *supra* note 66, at 214. For a current factual example, see Indictment at 6, U.S. v. Artem Radchenko and Oleksandr Ieremenko, 2018R01347 / DS (D.C. N.J.), https://www.justice.gov/usao-nj/press-release/file/1124251/download [https://perma.cc/W7GW-F5AN]. "It was part of the conspiracy that the defendants and others gained unauthorized access to the computer networks of the SEC by employing a variety of hacking methods, including directory traversal attacks and phishing attacks. The co-conspirators took steps to conceal and misrepresent their identities to illegally gain access to information on the internal networks of the SEC and to avoid detection." *Id.*

85. For a first approximation in the words of former SEC Commissioner Roberta S. Karmel, see Roberta S. Karmel *The Fiduciary Principle of Insider Trading Needs Revision*, 56 WASH. U. J. L. & POL'Y 121, 130 (2018). "[A] hacker is not anyone's fiduciary…." *Id.*

86. *See* hiQ Labs, Inc. v. LinkedIn Corp., No. 17-16783, 2019 WL 4251889, at 14 (9th Cir. Sept. 9, 2019) ("[I]t appears that the CFAA's prohibition on accessing a computer "without authorization" is violated when a person circumvents a computer's generally applicable rules regarding access permissions, such as username and password requirements, to gain access to a computer. It is likely that when a computer network generally permits public access to its data, a user's accessing that publicly available data will not constitute access without authorization under the CFAA. The data hiQ seeks to access is not owned by LinkedIn and has not been demarcated by LinkedIn as private using such an authorization system."); hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1113 ("hiQ's circumvention of LinkedIn's measures to prevent use of bots and implementation of IP address blocks does not violate the CFAA because hiQ accessed only publicly viewable data not protected by an authentication gateway"); Sandvig v. Sessions, 315 F. Supp. 3d 1, 27 (D.D.C. 2018) ("Employing a bot to crawl a website or apply for jobs may run afoul of a website's ToS, but it does not constitute an access violation when the human who creates the bot is otherwise allowed to read and interact with that site."). *See* Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058, 1068 n.5 (9th Cir. 2016) ("Simply bypassing an IP address, without more, would not constitute unauthorized use."); Ticketmaster L.L.C. v. Prestige Entm't W., Inc., 315 F. Supp. 3d 1147, 1171 (C.D. Cal. 2018) ("To be clear, it is the violation of the terms of the Letter, not of Ticketmaster's Terms of Use, on which the Court bases its

Congress – their a priori fiduciary relationship to the public at large being somewhat more easily justifiable due to an implicit function as trustees for the public good – the legislature saw the need to explicitly provide for such a fiduciary relationship via the STOCK Act of 2012[87] to combat the dangers of the burgeoning profession of lobbyism. Thus, even if the terms of use of a website bar scraping or any other commercial use of information presented on the site or its operating procedure tries to do so via technological means, it remains unclear whether courts would be willing to equate that with the existence of a fiduciary relationship between operator and user.[88] In answering this difficult question, one should rely on the basic rationale behind barring insider trading, which is to prevent fraud on the market[89] to preserve investors' faith in the institutional integrity of capital markets[90] – and such fraud is simply not at stake in this scenario. If a website operator decides to run a public forum for greater visibility and does not implement any authentication gateway or explicitly restrict the access of a user found to be in violation of his terms of use[91] – which would undoubtedly be both within his rights and his capabilities – accessing said content in any way, shape or form (i.e. even via a scraping algorithm outsmarting CAPTCHA-technology) simply does result in deceit, as it has to be within the expected (disapproved of, it may be) behavior of users.[92]

---

finding of a well-pled CFAA claim.").

87. Stop Trading on Congressional Knowledge (STOCK) Act of 2012, Pub. L. No. 112-105; 126 Stat. 291.

88. *See* S.E.C. v. Dorozhko, 574 F.3d 42, 51 (2d Cir. 2009) ("[U]nclear, however, that exploiting a weakness in an electronic code to gain unauthorized access is 'deceptive,' rather than being mere theft."); United States v. Nosal, 844 F.3d 1024, 1038 (9th Cir. 2016) (Refusing to equate a circumvention of a website's terms of use with a CFAA violation: "violating use restrictions, like a website's terms of use, is insufficient without more to form the basis for liability under the CFAA.").

89. *See* Basic Inc. v. Levinson, 485 U.S. 224, 250 (1988) ("It is not inappropriate to apply a presumption of reliance supported by the fraud-on-the-market theory.").

90. *See* Karmel, *supra* note 85, at 134 ("[T]he ban on insider trading is important to investor confidence in the markets.").

91. *See* Craigslist Inc. v. 3Taps Inc., 964 F. Supp. 2d 1178, 1184 (N.D. Cal. 2013) ("[The] average person does not use 'anonymous proxies' to bypass an IP block set up to enforce a banning communicated via personally-addressed cease-and-desist letter.").

92. For behavior that might rise to a fraudulent representation in the digital world, see BLUE RIVER PARTNERS, ANNUAL COMPLIANCE UPDATE 10 (2018), http://www.blueriver partn ersllc.com/wp-content/uploads/2018/01/2017-Year-End-Update-FINAL-01.08.18.pdf [https: //perma.cc/4Q4Y-NLTF]; "[A] potentially deceptive practice may include a data provider's use of web scraping to gather information while disregarding or evading a website's security protocols." *Id.*; Robert G. Leonard, Jeffrey D. Neuburger & Joshua M. Newville, *Best Practices for Private Fund Advisers to Manage the Risks of Big Data and Web Scraping*, HEDGE FUND L. REP. (June 15, 2017), https://www.hfla wreport.com/2552996/best-practices-for-private-fund-advisers-to-manage-the-risks-of-big-data-and-web-scraping.thtml [https://p

(c) Violation of Terms of Use as Deceptive Behavior?

The boilerplate acquiescence to contractual terms of use upon entering a site can and should not be equated with an affirmative representation that any and all such terms will be honored in the future, thereby exaggeratedly superimposing fiduciary breaches for the purposes of insider trading doctrine on any run-of-the-mill contract breach: Subsequent contractual breaches should not automatically imply fraudulent inducement in entering into an agreement in the first place.[93] Doing so would, viewed in a larger, doctrinal context, greatly endanger contractual stability as every agreement would be hanging in the balance as long as breaches are still possible, since such a breach would then serve as proof of earlier fraudulent inducement. The site proprietor may have tried to limit the re-use of the information he displayed but he never envisioned that certain individuals would be cut off entirely from his data,[94] i.e. he likely would not feel deceived through DataPortal accessing his site from the perspective of a reasonable bystander.[95] Policy

erma.cc/Y987-8HLY] ("Circumventing security protocols; disguising or failing to reveal a scraper's identity on a site (where required); and simulating human transactions, among other behaviors, each could be viewed as an affirmative misrepresentation").

93. *But see* Butterworth, *supra* note 68 ("[A] potentially deceptive tactic occurs when a webscraper, in exchange for access to website information, agrees to terms and conditions which prohibit scraping activities."); Sanaea Daruwell, *Navigating Compliance When Extracting Web Scraped Alternative Financial Data*, THE SCRAPINGHUB BLOG (Mar. 21, 2019), https://blog.scrapinghub.com/regulatory-compliance-for-alternative-web-scraped-financial-data [https://perma.cc/7PJ2-3DMA] ("If the terms state that you may not scrape the site or use automated means to extract data from the site, your web scraping project may not only [!] give rise to insider information issues, but also to breach of contract claims.").

94. *See* hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985, 1002 (9th Cir. 2019) ("[U]nderstanding that the CFAA is premised on a distinction between information presumptively accessible to the general public and information for which authorization is generally required").

95. *See* Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058, 1070 n.3 (9th Cir. 2016) ("Violation of Facebook's terms of use, without more, would not be sufficient to impose liability. Nosal I, 676 F.3d at 862–63. But, in addition to asserting a violation of Facebook's terms of use, the cease and desist letter warned Power that it may have violated federal and state law and plainly put Power on notice that it was no longer authorized to access Facebook's computers."); hiQ Labs, Inc. v. LinkedIn Corp., 273 F.Supp.3d 1099, 1113 n.9 ("[W]hen a business displays a sign in a storefront window for the public to view, it may not prohibit on pain of trespass a viewer from photographing that sign or viewing it with glare reducing sunglasses."); Sandvig v. Sessions, 315 F. Supp. 3d 1, 27 (D.D.C. 2018) ("The website might purport to be limiting the identities of those entitled to enter the site, so that humans but not robots can get in. See Star Wars: Episode IV—A New Hope (Lucasfilm 1977) ('We don't serve their kind here! . . . Your droids. They'll have to wait outside.'). But bots are simply technological tools for humans to more efficiently collect and process information that they could otherwise access manually. See Star Wars: Episode II—Attack of the Clones (Lucasfilm 2002) ('[I]f droids could think, there'd be none of us here, would there?').").

concerns regarding the looming ineffectiveness of a website's terms of use are to be heeded, but are misplaced within the insider trading doctrine: Enforcing terms of use of a public domain is not the job of Section 10(b) and Rule 10b-5, but rather a task to be brought about by ordinary contract and tort law.[96]

SEC v. Dot9: DataPortal scraped customerreviews.com in contravention of the website's terms of use. In addition, its scraping algorithm was able to circumvent the CAPTCHA-technology that was primarily designed to keep bots off of the webpage. Both may well be actionable from LocalCookBook's point of view with regard to DataPortal[97], but the disclosure of the scraped data set to Dot9 is not a violation of DataPortal's fiduciary duties towards LocalCookBook (if there even are any) as the access is not "unauthorized" under the CFAA and therefore likely not based on deceit under the insider trading doctrine.

96. *See* hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985, 1004 ("We note that entities that view themselves as victims of data scraping are not without resort, even if the CFAA does not apply: state law trespass to chattels claims may still be available. And other causes of action, such as copyright infringement, misappropriation, unjust enrichment, conversion, breach of contract, or breach of privacy, may also lie."); ebay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1070 (N.D. Cal. 2000) (equating scraping in contravention of a "robots.txt"-restriction with digital trespass with regard to the resulting server load); Sw. Airlines Co. v. Roundpipe, LLC, 375 F. Supp. 3d. 687, 706 (N.D. Tex. 2019) ("[The] court concludes that Southwest's complaint states a plausible claim for breach of contract because Southwest's complaint not only identifies the existence of a valid contract (Southwest's use agreement) but it also explains how the defendants' use of automated scraping tools breached the contract and caused damage to Southwest"). *See also* Joshua A. T. Fairfield, *"Do-Not-Track" as Contract*, 14 VAND. J. ENT. AND TECH. L. 545, 582 (2012) ("Likewise, when one website wishes to exclude robots and scrapers from its service, the website posts a file called "robots.txt" that includes the restrictions on scraping. These restrictions are readable by other people's scrapers and agents, and are quite binding: if the scraper continues despite the preferences expressed in the robots.txt file, courts have analogized the resulting server load to trespassing on someone's land without permission. Thus, automated contracts are enforced."); Hirschey, *supra* note 26, at 918 ("[A] data host should only exercise these legal options if scrapers seek to challenge the data host's business model parasitically and not to augment it mutualistically"); Leonard, Neuburger & Newville, *supra* note 92 ("Violation of the EULA [i.e. end-user license agreements] by, for example, scraping information has been used successfully as the basis for breach-of-contract claims."). *See* Horton, *supra* note 11, at 8 (noting that given the unlikelihood of pursuing these claims, violating the terms of use is a "business risk that a lot of managers may be willing to take").

97. *See* hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985, 1004 n.15 ("[I]t may be that web scraping exceeding the scope of the website owner's consent gives rise to a common law tort claim for trespass to chattels, at least when it causes demonstrable harm.").

4. Role of Consumer Consent to Gather Scraped Data

If the (contractual) terms with the customer allow for the commercial use of any such (anonymized) data – which may well be the case, given the relatively little attention that customers have paid to issues of data privacy in the past[98] – that data is often packaged and sold for either advertising purposes or other analytical objectives which also includes investment funds as possible buyers of such data. Regarding geolocation data, the origins of such data will likely lie with either telecommunications companies (e.g. Verizon, AT&T, T-Mobile), with the proprietors of cellular operating systems (e.g. Apple, Google[99]) or with any other company that gathers constant data streams on smartphones through single apps.

(a) Foreseeability of Use for the Consumer?

Looking at the true meaning of consent in more detail, the issue arises of whether the user merely needs to be notified of the possibility that his location data may be used or whether he needs to be notified of what exactly the gathering entity intends to do with it (e.g. sale of data to an investor). Oftentimes the disclosure of intent by the data gathering entity is very generic and may at no point indicate the actual intended commercial use[100]:

98. *See* CITI BUS. ADVISORY SERVS., *supra* note 59, at 17 (referring to a real-world example in which "semi-private" data is being used for investment purposes: "App firms that give away free services, such as email readers, are scrubbing their user base communications for confirmation emails on purchased items and providing these receipts with SKU-level data in bulk to investment managers on a monthly basis."); Fortado, *supra* note 58 (quoting a former prosecutor and current attorney at Dechert LLP: "You look in the small print and there's probably somewhere in there that says Verizon can sell that data.").

99. *See* Jennifer Valentino-DeVries, *Tracking Phones, Google Is a Dragnet for the Police*, N.Y. TIMES (April 13, 2019), https://www.nytimes.com/interactive/2019/04/13/us/google-location-tracking-police.html?action=click&module=Top%20Stories&pgtype=Hom ep age [https://perma.cc/EPT9-4YV9] ("Location data is a lucrative business — and Google is by far the biggest player, propelled largely by its Android phones. It uses the data to power advertising tailored to a person's location, part of a more than $20 billion market for location-based ads last year.").

100. *See* Valentino-DeVries et. al., *supra* note 28 (citing Emmett Kilduff, CEO of Eagle Alpha, a data broker for investment managers: "Most people don't know what's going on."). If the information is "material to consumers [acting reasonably under the circumstances] [. . .] [and] would likely affect the consumer's conduct or decision with regard to a product or service", it constitutes a deceptive business practice to be curbed by the FTC. FED. TRADE COMM'N, ENFORCEMENT POLICY STATEMENT ON DECEPTIVELY FORMATTED ADVERTISEMENTS 1 (Dec. 22, 2015), https://www.ftc.gov/system/files/documents/public_statements/896923/151222deceptiveenforcement.pdf [https://perma.cc/QL7D-W7WJ]. For more details on the "reasonable consumer"-standard, see Celine Shirooni, *Native Advertising in Social Media: Is*

Even the very liberal phrasing of the Terms of Use of The Weather Channel App that only refer to "analyzing trends based on foot traffic" as their example for "commercial purposes"[101] is not immediately suggestive of helping investors generate returns by privately using one's smartphone data.

(b) Possible Deception of the Consumer?

It would not be a stretch to state that activating the gathering of location data for a weather app would seem like a necessity for a useful weather app on a smartphone: How else will the app know which city's weather to show on the homescreen? If the user is expecting his consent for use of his data to be necessary for the service in question to perform as advertised, it is all the more unlikely that the customer will even take a closer look at the terms of such a consent. Therefore, the commercial sale of such data to third parties such as hedge funds for investment purposes that the customer himself neither (reasonably) foresees nor in any way profits from may be somewhat surprising.[102] As such, the collection and sale of data for investment purposes that rely only on a blanket consent of the user may violate federal laws[103] against deceptive business practices.[104] Whether that can be equated with the

the FTC's Reasonable Consumer Reasonable, 56 WASH. U. J. L. & POL'Y 221, 234–39 (2018).

101. *See* THE WEATHER CHANNEL, PRIVACY POLICY 3(F) (December 5, 2018), https://weather.com/en-US/twc/privacy-policy#us-how-we-share-new [https://perma.cc/U8PY-KF3Y] ("As part of the Services, we may aggregate or otherwise alter information (including location information) that is collected from the Services so that it does not identify your device and cannot reasonably be linked to your device. We may use or share such information with third parties for research or commercial purposes (e.g., analyzing trends based on foot traffic)."). This clarification was added in response to an inquiry from The New York Times to the parent company, IBM, in the course of an inquiry about business practices in commercial use of smartphone data. Valentino-DeVries et. al., *supra* note 28.

102. *But see* Enrique Dans, *They Wouldn't Sell Your Geolocation Data Without Your Permission . . .*, FORBES (Dec. 11, 2018), https://www.forbes.com/sites/enriquedans/ 2018/12/11/they-wouldnt-sell-your-geolocation-data-without-your-permission/#400da2c77f08 [https://perma.cc/Z56R-T7MT] ("The issue here is whether it is reasonable for an app to tuck away a clause in its terms of service that allows it to sell its users' sensitive geolocation data. The answer is obvious: any app that does so should be fined heavily. Who would imagine that it was okay to sell highly sensitive geolocation data? In other words, if companies are doing so, it's clearly without users' knowledge.").

103. More specifically, see Section 5 of the FTC Act (15 U.S.C. § 45(a)(1) (2006)) which prohibits "unfair or deceptive acts or practices in or affecting commerce".

104. *See* Valentino-DeVries et. al., *supra* note 28 (citing Maneesha Mithal of the FTC: "You can't cure a misleading just-in-time disclosure with information in a privacy policy."). The FTC recently reached a settlement with PayPal over allegations that the company was misleading Venmo customers "about the extent to which they could control the privacy of their transactions." FED. TRADE COMM'N, PAYPAL SETTLES FTC CHARGES THAT VENMO FAILED TO DISCLOSE INFORMATION TO CONSUMERS ABOUT THE ABILITY TO TRANSFER FUNDS

deception required for a breach of duty under insider trading law is unclear[105]: While a deceptive business practice innately contains an element of deceit – and, as such, arguably contains fraudulent behavior by definition –, it is likely not exactly of the same nature as the concretely fraudulent act that is required for insider trading under Section 10(b) and Rule 10b–5. A misleading business practice may be contravening trade and consumer protection regulation but does not necessarily rise to the level of affirmative misrepresentation that is required when access to digital information is concerned.[106]

(c) Impact on Consumer Consent

Unless there is fraud with respect to the relevant part of the terms that allow for a gathering of data that is later used as part of the investment research, that customer consent will likely be enough for hedge funds trading on the gathered data to avoid insider trading liability as there is not even a contractual breach.

> SEC v. Dot9: The location data that DataPortal used as the basis for its analysis could be gathered and sold in accordance with the terms of use of LocalCookBook. As there is no clear indication of fraud – in particular, there is likely no reliance[107] of the customer on the fact that the data he is explicitly allowing the app proprietor to gather would not be used for commercial purposes such as investment research (especially considering the ubiquity of user preferences-based advertising that obviously rests on the commercial use of customer data) – that consent is

AND PRIVACY SETTINGS; VIOLATED GRAMM-LEACH-BLILEY ACT (Feb. 27, 2018), https://www.ftc.gov/news-events/press-releases/2018/02/paypal-settles-ftc-charges-venmo-failed-disclose-information [https://perma.cc/F9PF-2TKN].

105. *See* THE ECONOMIST, *supra* note 4 ("Since many phone and credit-card companies include clauses in their contracts allowing them to sell information, that condition [i.e. "breach of duty"] is rarely fulfilled.").

106. *See* United States v. Finnerty, 533 F.3d 143, 148 (2d Cir. 2008) ("Broad as the concept of "deception" may be, it irreducibly entails some act that gives the victim a false impression."); S.E.C. v. Dorozhko, 574 F.3d 42, 51 (2d Cir. 2009) ("[M]isrepresenting one's identity in order to gain access to information that is otherwise off limits, and then stealing that information is plainly 'deceptive' within the ordinary meaning of the word.").

107. *See* Basic Inc. v. Levinson, 485 U.S. 224, 243 (1988) ("[R]eliance is and long has been an element of common-law fraud, see, e.g., Restatement (Second) of Torts § 525 (1977); W. Keeton, D. Dobbs, R. Keeton, & D. Owen, Prosser and Keeton on Law of Torts § 108 (5th ed. 1984).").

> likely to be considered valid and encompasses the sale of data to DataPortal. As there was no breach of fiduciary duty through the sale of the customer data from LocalCookBook to DataPortal, there can be no derivative breach in DataPortal's subsequent sale of that data to Dot9, equally ruling out an insider trading violation by Dot9.

When it all comes down to it, oftentimes the sale of user data to analytics firms or directly to hedge funds will be contractually allowed, no matter how obscure the clause may be[108] – and therefore not pose a scenario in which a breach of fiduciary duty is likely to occur.[109]

D. Personal Benefit to the Tipper / Tippee Liability / Scienter

It has been the subject of much discussion whether and to what extent a personal benefit must accrue to the insider when he relays the information to a third party in breach of his fiduciary duties. The current state of play is that there must a personal benefit conferred upon the insider.[110] This personal benefit does not, however, have to be pecuniary but can also lie in the immaterial pleasure taken in wanting to benefit a relative or a friend with a one-sided gift.[111] Some uncertainty remains whether gifting a stranger a

108. *See* Theodore Rostow, *What Happens When an Acquaintance Buys Your Data?: A New Privacy Harm in the Age of Data Brokers*, 34 YALE J. ON REG. 667, 706 (2017) (urging the FTC to require "logistically smooth opt-in provisions regarding the sale of collected data", implicitly acknowledging the reality that current practices, although questionable from a consumer perspective, are in line with current consent requirements).

109. Moral outrage does not negate contractual consent as long as neither fraud nor duress is present (only the former could possibly be relevant, but fraud seems like a severe stretch as long as one does not demand an explicit and itemized warning by the proprietor that any gathered data may be sold for investment research purposes). *See* THE ECONOMIST, *supra* note 4 ("[The] condition [breach of fiduciary duty] is rarely fulfilled"). For a different (emotionally charged) perspective see Dans, *supra* note 102. "Contrary to common belief, just because something is included in a contract and we sign it doesn't make it legal." *Id.*

110. *See* Dirks v. SEC., 463 U.S. 646, 662 (1983) ("[T]he test is whether the insider personally will benefit, directly or indirectly, from his disclosure"). *See also* Karmel, *supra* note 85, at 125 (being critical, both with an eye towards the judicial doctrine as well as towards the SEC's application of said doctrine: "Surely, it would have been better to develop a doctrine to distinguish between diligent research and information obtained through dishonest methods. On the other hand, the SEC should have developed such a doctrine rather than trying to cast the widest possible net to catch insider traders.").

111. *See* Salman v. United States, 137 S. Ct. 420, 428 (2016) (overruling United States v. Newman, 773 F.3d 438, 452 (2d Cir. 2014)) ("To the extent the Second Circuit held that the tipper must also receive something of a "pecuniary or similarly valuable nature" in exchange

piece of inside information also satisfies the requirement as the holding of the leading case on this area was somewhat narrow,[112] but in the context of alternative data accumulation by investment managers this is unlikely to be relevant as the data is all but certain to be paid for.

> SEC v. Dot9 In the case at hand, the prerequisite of DataPortal receiving a personal benefit for any disclosure is undoubtedly met through the monetary compensation DataPortal received from Dot9.[113]

In general, insider trading liability depends on an element of scienter of the trader regarding the breach of fiduciary duty. As a subjective element, scienter is understood to be "a mental state embracing intent to deceive, manipulate, or defraud."[114] If – as in the example case of SEC v. Dot9 – the

---

for a gift to family or friends, *Newman*, 773 F.3d, at 452, we agree with the Ninth Circuit that this requirement is inconsistent with *Dirks*."); United States v. Salman, 792 F.3d 1087, 1094 (9th Cir. 2015), *aff'd*, 137 S. Ct. 420 (2016) ("Proof that the insider disclosed material nonpublic information with the intent to benefit a trading relative or friend is sufficient to establish the breach of fiduciary duty element of insider trading."); *see also* Peter J. Henning, *Making Up Insider Trading Law as You Go*, 56 WASH. U. J. L. & POL'Y 101, 118 (2018) ("[Salman] turns out to be an uninteresting case that adds little to the law of insider trading.").

112. The U.S. Supreme Court likely did not want to equate the wilful act of information disclosure with automatically receiving a personal benefit (this would arguably have been sensible as voluntarily giving up something of value confers upon the grantor the benefit of allocating his resources according to his preferences which are paramount from an economical viewpoint). Matthew J. Wilkins, *You Don't Need Love . . . But It Helps: Insider Trading Law after* Salman, 106 KY. L.J. 433, 448 (2017). Otherwise its emphasis on potential difficulties for the courts with finding a personal (non-monetary) benefit in cases without relatives or friends going forward would be somewhat perplexing. *See* Salman v. United States, 137 S. Ct. 420, 429 (partially quoting Dirks v. SEC, 463 U.S. 646, 664 (1983)) ("It remains the case that '[d]etermining whether an insider personally benefits from a particular disclosure, a question of fact, will not always be easy for courts.' 463 U.S., at 664, 103 S.Ct. 3255. But there is no need for us to address those difficult cases today, because this case involves 'precisely the "gift of confidential information to a trading relative" that *Dirks* envisioned.'"). A newer case out of the Second Circuit, which previously narrowed the definition of "personal benefit" in *Newman*, argues now – in partial contradiction to *Newman* – that "sheer speculation into the tipper's motives" may actually give rise to a personal benefit even outside of relatives and friends as recipients when "circumstantial evidence" support the claim that the tipper wanted to benefit the tippee. United States v. Martoma, 894 F.3d 64, 76 (2d Cir. 2017) ("[T]he personal benefit element can be met by evidence that the tipper's disclosure of inside information was intended to benefit the tippee.").

113. Dirks v. SEC., 463 U.S. 646, 663 (1983) (finding that deciding whether there has been a breach of duty by the insider "requires courts to focus on objective criteria, i.e., whether the insider receives a direct or indirect personal benefit from the disclosure, such as a pecuniary gain").

114. *See* Ernst & Ernst v. Hochfelder, 425 U.S. 185, 193 n.12, 212–14 (1976) ("[H]istory

trader is relying on information furnished to him from a tipper, the disclosure from tipper to tippee itself must be a violation of a fiduciary duty as "absent a breach by the insider, there is no derivative breach [of the tippee]."[115]

A key element of tippee liability then is whether he knows or has reason to know that the disclosure of information to him constitutes a breach of fiduciary duty, thereby participating in the original breach through the use of the fiduciaries' information in securities trades.[116] Generally, the knowledge requirement is regarded as a meaningful barrier to the insider trading liability of hedge funds relying on information gathered, arranged and provided to them by third parties.[117] This holds true especially when the data provider stipulates to having obtained any and all furnished data lawfully and without improprieties, providing the hedge fund with a potential safe harbor excuse for allegations of insider trading.[118] Neither such warranties nor the knowledge requirement per se can, however, insulate hedge funds entirely from possible misconduct.

> SEC v. Dot9: The manager at Dot9 had knowledge of where the internal Apple customer data came from. He needn't have known the specifics of the contract between DataPortal and Apple to know (or reasonably infer) that disclosing such data to a third person – even integrated as part of an aggregation – would constitute a violation of DataPortal's fiduciary duties to its client. The fact that the data set contained Apple store customer analytics that could not possibly have originated from a public source

---

mak[es] clear that when the Commission adopted the Rule [i.e. Rule 10b-5] it was intended to apply only to activities that involved scienter."); *see also* Joan MacLeod Heminway, *Tipper/Tippee Insider Trading as Unlawful Deceptive Conduct: Insider Gifts of Material Nonpublic Information to Strangers*, 56 WASH. U. J. L. & POL'Y 65, 91 (2018) ("No opinion of the U.S. Supreme Court has fully or formally established the nature of the required state of mind for insider trading liability . . . .").

115. Dirks v. SEC, 463 U.S. 646, 662 (1983); *see also* SEC v. Obus, 693 F.3d 276, 287 (2d Cir. 2012) ("A tippee must have some level of knowledge that by trading on the information the tippee is a participant in the tipper's breach of fiduciary duty.").

116. *See* Dirks, 463 U.S. at 660 n.19 (quoting Inv'rs Mgmt. Co., Inc., 44 S.E.C. 633 (July 29, 1971)) ("[O]ne element of tippee liability is that the tippee knew or had reason to know that the information 'was non-public and had been obtained improperly by selective revelation or otherwise.'").

117. For hedge funds' possible insider trading liability when relying on acquired political intelligence, see Pauley, *supra* note 2, at 645. "[F]inding knowledge is a high hurdle." *Id.*

118. *See* Pauley, *supra* note 2, at 646 (explaining that PI firms' statements that they have disavowed insider tactics provides hedge funds with a safe harbour to argue that they did not know that information was obtained illegally).

> (who else could gather concrete data on which portion of registered Apple store customers complete a purchase other than Apple itself or an agent thereof?) should have raised an immediate red flag with a sophisticated investor.[119] The assurance by DataPortal that all information provided had been obtained legally, is of no effect in the face of actual knowledge of a fiduciary breach.[120] By at least consciously disregarding the flaring warning signal, the PM at Dot9 acted therefore with scienter and is likely liable for an insider trading violation.

E. Interim Finding

Even if cases in the digital world do exist in which insider trading liability comes into play[121] – especially when hacking is involved or proprietary data is transmitted in violation of contractual and fiduciary duties –, scraped data generally does not give rise to criminal or civil liability under Section 10(b) and Rule 10b-5.

> SEC v. Dot9: Only the information regarding the buying/grievance-airing customer split that DataPortal misappropriated through breaching its fiduciary duty to its client, Apple, will likely have given rise to an insider

119. *See* SEC. v. Obus, 693 F.3d 276, 288 (2d Cir. 2012) ("[This is a] fact-specific inquiry turning on the tippee's own knowledge and sophistication and on whether the tipper's conduct raised red flags that confidential information was being transmitted improperly.").

120. *See* United States v. Martoma, 869 F.3d 58, 61–62 (2017) (declaring the consultant's assurance "not to disclose any confidential information in a consultation" to be without effect on the defendant's guilt).

121. In rare cases, the breach of some sort of fiduciary duty may seem obvious. *See* THE ECONOMIST, *supra* note 4 (describing a case in which an ex-employee of the U.S. federal government reportedly offered "predictions" of reports his former agency was working on for sale to a hedge fund; if such "predictions" are to carry any value, some fiduciary duty breach stemming from the relationship agency-(ex-)employee seems more than likely). Similarly, the SEC opened an administrative proceedings against Deerfield, an investment adviser in the health care sector, and subsequently settled for a penalty of around $4 million, under the following circumstances. *See* Deerfield Management Company, L.P., SEC Docket (CCH) 3608571 (2017) ("In May and June 2012, the political intelligence analyst provided Deerfield analysts with specific information regarding confidential CMS deliberations regarding cuts to Medicare reimbursement rates for certain radiation oncology treatments. The Deerfield analysts recommended trades based on this information, and Deerfield then traded on behalf of certain of its hedge funds to sell short shares of two companies who offered products and services related to radiation oncology. The hedge funds that Deerfield advised then profited when CMS announced the rate cuts.").

trading liability of Dot9's PM.[122]

While that should put hedge funds on guard against analytics companies pitching them scraped datasets that noticeably implicate proprietary information or information that can't possibly be public and should also motivate them to implement a judicious screening process[123] before data sets are purchased or internal investment research utilizing such data sets is part of the mix of information underpinning a trade,[124] the scienter requirement and the prerequisite that there must be a breach of fiduciary responsibility (or a fraudulent equivalent thereof) in the acquisition and/or the disclosure of the information at hand act as sensible restrictions on insider trading doctrine in the digital world.

122. *See* BLUE RIVER PARTNERS, *supra* note 92, at 10 ("[I]f the adviser's receipt of scraped data results in a violation of confidentiality obligations, and the adviser uses that data for securities transactions, it may risk the adviser violating anti-fraud provisions pursuant to the misappropriation theory.").

123. *See* Altman et. al., *supra* note 13, at 4 ("[D]iligence includes determining who owns the data the firm is purchasing, and verifying that its vendors have the right to sell that data to the firm for the firm's intended use."); CITI BUS. ADVISORY SERVS., *supra* note 59, at 17 (adding the following caveat to the utility of alternative data for investment strategies: "where there is no prohibition against mining obtainable data"); Sam Dale, *Alternative Data: What are the Regulatory Risks?*, HFMWEEK 16, 17 (Nov. 9–15, 2017) (referring to the common practice of a general counsel at a $2 billion hedge fund: "insert[ing] a covenant in our agreement to say they [i.e. the alternative data provider] won't provide us with MNPI or any information in breach of a duty to any third party"); Horton, *supra* note 11, at 7 ("[Techniques include] adopting a policy regarding insider trading; recordkeeping; implementing employee training programs; monitoring employees' personal securities trading; maintaining information barriers; and enforcing issues internally"); Pauley, *supra* note 2, at 646 ("[I]t would be wise for hedge funds . . . to limit their communications . . . to only the final product, which ideally would contain both a disclaimer and a statement of sources and methods."); Butterworth, *supra* note 68 ("Investors using web scraping techniques should therefore carefully consider any processes used and assess whether they involve elements of deceit and identity masking."); Tom Hardin, *The Inside Scoop on Insider Trading Prevention: Best Practices for Hedge Funds*, NICE ACTIMIZE (Oct. 5, 2017), https://www.niceactimize.com/blog/The-Inside-Scoop-on-Insider-Trading-Prevention-Best-Practices-for-Hedge-Funds-557 [https://perma.cc/GG52-LLPG] ("[C]ompliance is frequently under pressure from analysts to permit a data source on the basis that other firms are using that source.").

124. Actual reliance on material, non-public information for a specific trade is not required for insider trading liability. *See* United States v. Rajaratnam, 719 F.3d 139, 159–60 (2d Cir. 2013) (holding that possession of such information while executing a trade is sufficient).

IV. NORMATIVE ANALYSIS – ARE THERE LEGITIMATE REASONS TO SUBJUGATE TRADES BASED ON SCRAPED DATA TO INSIDER TRADING DOCTRINE?

As recent years have shown, insider trading law is a malleable instrument that can be employed in varying sets of circumstances to disincentivize harmful behavior.[125] The history of insider trading law itself shows that its broad and potentially all-encompassing wording was intended to provide flexibility for the law to evolve in the face of "cunning devices."[126] Throughout the twenty-first century, nothing has proven more cunning than the technological capacities of the ever-evolving IT world. The relatively new phenomenon of data scraping then seems primed to be – undeservedly, perhaps – understood as a threatening new technology that deepens or even creates information asymmetries incompatible with the concept of a free market. While technological progress is usually outwardly applauded, those same newly created possibilities can quickly become the reason for crying foul in the face of unruly capital markets gains, potential unfairness being one of the driving forces behind advancements in insider trading regulation[127].

---

125. *See* Henning, *supra* note 111, at 120 ("[Insider trading is] an area of the law that is the product of judicial creation, with a little help from the SEC in its rulemaking"); Reed Harasimowicz, *Nothing New, Man!-The Second Circuit's Clarification of Insider Trading Liability in* United States v. Newman *Comes at A Critical Juncture in the Evolution of Insider Trading*, 57 B.C. L. REV. 765, 800 (2016) ("The law of insider trading remains largely convoluted, especially with attempts by the SEC to expand the boundaries of the law with novel prosecution strategies against remote tippees."); Yesha Yadav, *Insider Information and the Limits of Insider Trading*, 56 WASH. U. J. L. & POL'Y 135, 141 (2018) (explaining that "[c]ourts have shown themselves willing to stretch legal interpretation to find that a breach of a fiduciary duty has taken place in order to impose sanctions on bad actors," giving the liability for hackers in *Dorozhko* as a prime example of this functional application of insider trading doctrine).

126. Sec. & Exch. Comm'n v. Texas Gulf Sulphur Co., 401 F.2d 833, 884 (2d Cir. 1968) (quoting *Stock Exchange Regulation: Hearing on H.R. 7852 and H.R. 8720 Before the H. Comm. on Interstate and Foreign Commerce*, 73d Cong. 115 [1934] [statement of Thomas Corcoran, Counsel with the Reconstruction Finance Corporation]); *see also* Zachary J. Gubler, "*Maximalism with an Experimental Twist": Insider Trading Law at the Supreme Court*, 56 WASH. U. J. L. & POL'Y 49, 55 (2018) ("Congress seems to have implicitly delegated lawmaking authority on insider trading law to the Court . . . .").

127. *See* Gina-Gail S. Fletcher, *Legitimate yet Manipulative: The Conundrum of Open-Market Manipulation*, 68 DUKE L. J. 479, 531 (2018) ("[T]he grounds for punishing insider trading . . . are often phrased in terms of fairness . . . ."); Robert A. Prentice, *The Internet and Its Challenges for the Future of Insider Trading Regulation*, 12 HARV. J. L. & TECH. 263, 305 (1999) (rightfully equating the perceived unfairness of an activity with the resulting thrust behind any regulatory activity); s*ee also* James Walsh, *"Look Then to Be Well Edified, When the Fool Delivers the Madman": Insider-Trading Regulation After* Salman v United States,

In line with that suspicion, while perusing the broad and multi-faceted range of literature on this particular topic, one can quickly get the impression that insider trading doctrine might either already be applicable to scraped data or commentators are assuming that it might be in the future.[128] Even if the current case law does not seem to put trading on scraped data within the ambit of insider trading doctrine, one might therefore wonder whether this, asking from a purely normative standpoint, might be the case in the future.[129] This thought experiment, which is supposed to shed light on the motivation behind imposing insider trading restrictions on scraped data as well as

67 CASE W. RES. L. REV. 979, 996 (2017) ("[A]s a public policy matter, the time has come for the federal government to take a hardline stance on insider trading, because there's no telling how many more *Newman's* are waiting in the wings of appellate courts, and how many insiders are out there seeking "recognition" and industry status as players with reliable information.").

128. *See* BLUE RIVER PARTNERS, *supra* note 92, at 10 ("Advisers engaging in web scraping must be aware of regulatory risks stemming from MNPI and insider trading concerns."); CITI BUS. ADVISORY SERVS., *supra* note 59, at 17 ("[I]nvestment managers will have to walk a fine line and make sure that that they are complying with insider trading regulations . . . ."); Dale, *supra* note 123, at 16 (quoting Doug Dannemiller, investment research leader for the Deloitte Center for Financial Services) ("There is a lot of undefined space with alterative data, but venturing into material non-public information (MNPI) is a main concern . . . ."); DELOITTE CTR. FOR FIN. SERVS., *supra* note 3, at 6–7, 15 ("Material nonpublic information (MNPI) risk: . . . [I]f an alternative data set is thought to be too predictive of normally protected information such as quarterly revenue, then some firms are steering clear of the data . . . . There are still open questions about acceptable practices in the areas of web-gathered information . . . . [T]here are certainly risks associated with incorporating alternative data into investment-decision processes . . . ."); INTEGRITY RESEARCH ASSOCS., *supra* note 38, at 8 (rating the insider trading risk for the alternative data sources of credit card transactions, email receipts, geolocation and satellites – all of which can "possibl[y]" be acquired on an exclusive basis – as "[h]igher"); Crane, Crotty & Umar, *supra* note 3, at 3 n.3 ("Note that private information is not necessarily illegal insider information but could instead stem from hard-to-get or costly data sources (e.g., satellite data, mobile phone data) . . . ."); THE ECONOMIST, *supra* note 4 ("[T]he need to sort useful from pointless, and legal from dubious, has never been greater."); Wieczner, *supra* note 3 ("Web scraping and other alternative data collection practices are already fueling debate over what constitutes nonpublic information and insider trading—and whether investors can misuse information even when it's public and legally obtained."); Butterworth, *supra* note 68 ("[R]isk of liability under Section 10(b) of the Exchange Act and SEC Rule 10b-5 is particularly sharp regarding the use of web-scraped data."); Daruwell, *supra* note 93 ("One of the larger risks associated with the use of data extracted from the web for investment decision making is the risk of obtaining insider information . . . ."); Hardin, *supra* note 123 ("Today, there are a lot of potential issues with respect to data sets and data scraping, where firms are coming into possession of semi-private data in the research process.").

129. *See* Dale, *supra* note 123, at 17 ("Hedge fund compliance professionals and lawyers say alternative data is not currently a top priority for the SEC, but that could change. [. . .] Regulations are not evolving as quickly as the technology but this could change very soon").

indicate the likelihood of such a development[130] — there being potentially huge implications for the research operations of investment managers — is guided by privacy concerns, the possible impact on competition, potential skewing of market incentives and, perhaps most importantly, the compatibility of such an application with the underlying doctrine of insider trading law.

A. Privacy and Data Protection Concerns

On the one hand, encouraging investors to gather and dissect data by all means necessary to stay ahead of the curve arguably creates skewed incentives: When data scraping involves, for example, violating individual's privacy rights, the incentive-structure of allowing this practice to go forward (generated returns not subject to disgorgement) may result in a negative, external effect.[131] While the benefits of trading on the scraped data are captured only by the employees and shareholders of the respective investor, the fall-out regarding the further erosion of privacy (e.g. economically relevant, increased reluctance of people to share data or partake in an offering by a company; loss of the public trust in corporate data protection promises that may lead to underinvestment and subsequent, avoidable insolvencies) has to be shouldered by society writ large. In addition, terms of use regarding web-based scraping — even if their content may be questionable from an economical standpoint — would be rendered partially ineffective by allowing illegally scraped data sets to be retailed without fear of insider trading doctrine kicking in, potentially lowering trust in contractual freedom and the legal system as a whole. All that could be avoided by integrating (perceived) privacy concerns into insider trading doctrine in a way that removes the incentive for data harvesting in an area (e.g. location data, medical records) where privacy violations are likely to arise.[132]

---

130. *See also* Ronald J. Colombo, *Buy, Sell, or Hold? Analyst Fraud from Economic and Natural Law Perspectives*, 73 BROOK. L. REV. 91, 118 (2007) ("But this begs an interesting question: is Rule 10b-5's apparent inability truly a shortcoming? Perhaps the failure of the traditional elements of Rule 10b-5 to cover the phenomenon of analyst fraud suggests that such fraud ought not be subject to sanction. [. . .] For these and similar questions, a normative lens is needed through which securities law and policy can be analyzed.").

131. This is a term stemming from microeconomics. S*ee* DAVID H. HYMAN, PUBLIC FINANCE 99 (9th ed. 2008) ("When a negative externality exists, the price of a good or service does not reflect the full marginal cost of resources allocated to its production").

132. This would prevent data brokers from shunning responsibility for previously committed privacy violations. *See* Valentino-DeVries et. al., *supra* note 28 ("'Most people don't know what's going on,' said Emmett Kilduff, the chief executive of Eagle Alpha, which

On the other hand, while some of the privacy, data protection and competition concerns that engulf the widespread practice of data scraping may seem legitimate,[133] insider trading law seems like the wrong instrument to combat those.[134] Privacy law and its enforcement should be vigorously strengthened to address any problems identified in those areas.[135] Outlawing investment strategies that are based on scraping only removes one of a host of incentives to commit the perceived privacy violations. Data protection

---

sells data to financial firms and hedge funds. Mr. Kilduff said responsibility for complying with data-gathering regulations fell to the companies that collected it from people.").

133. *But see* Rostow, *supra* note 108, at 678 ("As a general rule, statutes do not prevent brokers from buying and selling an enormous amount of information, digitally produced by consumers, relating to their health and physiology, cognitive abilities, interests, purchases, wealth, compulsions and social networks."). For general commenting on possible privacy erosions through the rise of data hegemons, see Maurice E. Stucke, *Should We Be Concerned About Data-Opolies?*, 2 GEO. L. TECH. REV. 275, 290 (2018). "Network effects and other entry barriers protect data-opolies from many forms of competition. As a result, they can depress an important parameter of non-price competition, privacy protection, below competitive levels and collect personal data above competitive levels." *Id.*

134. *See* Bertrand Guerin & Anna Wolf-Posch, *Special Report of the German Monopolies Commission: Can Competition Law Address Challenges Raised by Digital Markets?*, 7 J. EUR. COMPETITION L. & PRAC. 30, 36 (2016) ("The Special Report recognises that the practices of search engines that are capable of raising competition concerns [i.e. scraping] can fully be addressed under existing competition law."); Stucke, *supra* note 133, at 323 (offering a solution to combating the dangers data-opolies pose to privacy expectations and a free market equilibrium: "[a]ntitrust enforcers must coordinate with privacy and consumer protection officials"). For a (very opinionated) motivation to regulate scraping that views scraping suspect, but is unlikely to be stemmed by the application of insider trading doctrine, see Robert Ridless, *This LinkedIn Lawsuit Proves The Left Doesn't Really Care About Securing Your Data*, THE FEDERALIST (Apr. 10, 2018), https://thefederalist.com/2018/04/10/linkedin-lawsuit-proves-left-doesnt-really-care-securing-data/ [https://perma.cc/P8EX-HL7R]. "As a society we are legitimately wondering whether informed user consent can really protect us. But what is more worrisome is that here, if LinkedIn loses, user consent will have ceased altogether to be a benchmark of best practices. This is far from reassuring. Even with more privacy regulation on the horizon, there is no guaranteeing the "intelligence" gathered on us is information we'd ever knowingly entrust to the often partisan actors who paternalistically claim the right to control it." *Id.*

135. *See, e.g.*, David Dayen, *Big Tech: The New Predatory Capitalism*, THE AMERICAN PROSPECT (Dec. 26, 2017), https://prospect.org/health/big-tech-new-predatory-capitalism/ [https://perma.cc/7WJP-A7KG] ("[L]egitimate fear that GDPR will threaten the data-profiling gravy train."); Sallie Ann Keller, Stephanie Shipp & Aaron Schroeder, *Does Big Data Change the Privacy Landscape? A Review of the Issues*, 3 ANNU. REV. STAT. APPL. 161, 162 (2016) ("The all data revolution is changing the focus of the privacy discussion from the masking and suppression of data in order to maintain confidentiality, to trust, policy, and governance."); Matsakis, *supra* note 28 ("Tech companies are also beginning to acknowledge that personal data collection needs to be regulated."); Rostow, *supra* note 108, at 706 ("[T]he FTC should explore requiring companies to include a narrow set of clear, logistically smooth opt-in [NB: as opposed to the currently widespread opt-out practice] provisions regarding the sale of collected data.").

laws—lying at the root of any perceived problems—should be the tool that advances individual citizens' interests in this regard.[136] A convincing argument can be made that insider trading law — which is based on the notion that trading gains should not stem from a fraudulently obtained edge in key information unknown to the market at large – is ill-fitting for this purpose: The goal of Section 10(b) and Rule 10b-5 is to prevent fraudulent trading,[137] i.e. trading where the information that is being depended upon is sourced through deceiving someone in violation of a fiduciary duty and subsequent reliance on that information on the trading market therefore reinforces that fraud as the trader in question is one step ahead of his fellow market participants in a direction in which other investors, even with smart and resourceful behavior, cannot possibly follow.[138] No matter the policy concerns with data scraping, that description, as it stands, simply does not fit the bill with regard to trading on alternative data sets. Moreover, by trying to patch up privacy law deficits with the hammer of insider trading liability, data protection laws will remain woefully inadequate in all other situations, i.e. where the dubiously acquired information is, for example, used for marketing purposes or as social leverage.[139] Calling for insider trading liability on scraped data sets may, with regard to consumer protection, hence lead to a pyrrhic victory.

B. Impact on Competition

If trading on scraped data is illegal, investors will have no incentive to purchase any data sets from data vendors. Data vendors that rely on innovative algorithms to scrape public information and other companies' websites (an emerging business model for newly conceived ventures) economically subsist on the interest of professional investors in their packaged product. Their entire business model depends on the ability of

136. *See* Matsakis, *supra* note 28 ("Even in a divided Congress, lawmakers could come together around privacy—scrutinizing Big Tech has become an important issue for both sides. . . . Until consumers actually understand the ecosystem they've unwittingly become a part of, we won't be able to grapple with it in the first place.").

137. See, for the Rule's substantive origin, Milton V. Freeman, *Administrative Procedures*, 22 BUS. LAW. 891, 922 (1967). "All the Commissioners read the rule and they tossed it on the table, indicating approval. Nobody said anything except Sumner Pike who said, "Well" he said, "we are against fraud, aren't we?" That is how it happened." *Id.*

138. For the SEC's view on this in Matter of Merrill Lynch, see Pierce, Fenner & Smith, Inc., 43 S.E.C. 933, 936 (Nov. 25, 1968). "[I]nherent unfairness involved where one takes advantage of such information knowing it is unavailable to the investing public." *Id.*

139. *But see* Rostow, *supra* note 108, at 673 (pointing to the danger of "relational control", when individuals are able to obtain telling information on one another via data brokers).

investors to use that information legally. All that will remain as a source for data that may lend itself to structuring a trading strategy around are digital behemoths that already have a significant portion of data under their (proprietary) control, giving them even an incentive to take over the business model of the data vendors.[140] In practically outlawing the use of scraped data for trading purposes one may unwillingly cement the market leader position of already uber-competitive data giants.[141] New ventures relying on novel data accumulation schemas that may challenge the hegemony of modern-day oil giants[142] will thusly be made more difficult to sustain.[143] That may be especially disheartening given the sheer ubiquity of ever-available data as the valuable skill nowadays should lie in developing a smart schema to gather and analyze information rather than in being able to successfully block off innovative search algorithms.[144] Those consequences seem

140. *See* hiQ Labs, Inc. v. LinkedIn Corp., No. 17-16783, 2019 WL 4251889, at 15 (9th Cir. Sept. 9, 2019) ("[G]iving companies like LinkedIn free rein to decide, on any basis, who can collect and use data—data that the companies do not own, that they otherwise make publicly available to viewers, and that the companies themselves collect and use—risks the possible creation of information monopolies that would disserve the public interest."); hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1117 (quoting from the hiQ brief: "LinkedIn is unfairly leveraging its power in the professional networking market to secure an anticompetitive advantage in another market—the data analytics market."); Eran Halevy, *Once Only for Huge Companies, 'Web Scraping' Is Now an Online Arms Race No Internet Marketer Can Avoid*, ENTREPRENEUR (Apr. 20, 2018), https://www.entrepreneur.com/article/311261 [https://perma.cc/UA24-U97X] ("What started as a one-way tool to extract web data and increase competition for the benefit of consumers turned into an arms race in which the target websites try to sabotage the data collection in order to achieve a competitive advantage.").

141. *See* hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1119 ("[W]ould delegate to private parties the sole authority to decide who gets to participate in the marketplace of ideas located in the "modern public square" of the Internet."). *See also* OECD, BIG DATA: BRINGING COMPETITION POLICY TO THE DIGITAL ERA 4 (Apr. 26, 2017), http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DAF/COMP/M%282016%292/ANN4/FINAL&docLanguage=En [https://perma.cc/J956-F6SY] ("A first approach to incorporate Big Data into competition law enforcement is to treat data as an input or asset that companies may use to enhance their market power and engage in exclusionary practices.")

142. *See* Matsakis, *supra* note 28 ("Personal data is often compared to oil—it powers today's most profitable corporations, just like fossil fuels energized those of the past.").

143. *See* Greene, *supra* note 83 (drawing a comparison to the hiQ/LinkedIn case: "Basically LinkedIn believes it has a right to deny 3rd parties, like HiQ, from scraping data which was intentionally set by users to be made publicly available.").

144. *See* hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1119 ("In view of the vast amount of information publicly available, the value and utility of much of that information is derived from the ability to find, aggregate, organize, and analyze data."). *See also* Halevy, *supra* note 140 ("More cunning is serving falsified information -- serving bots a higher-than-actual price, for example -- to foil the scraper's plan, rather than the mechanism."); Hirschey, *supra* note 26, at 926 ("[D]ata hosts that accept the valuable role of

unwelcome from an anti-trust perspective.[145] Furthermore, they might also forebode a future in which the free flow of information is negatively impacted:[146] Essentially, one might say in hyperbole, the free market, as reformed by the invention of the internet and the rise of data as the new gold, itself is at stake when the operators of websites that publicize information are able to stave off any professional analyzing of their data on display as it forestalls the dissemination and commercial use of material that should belong to the public sphere.[147] Making access to data more costly[148] or

scrapers in the digital environment stand to benefit from cooperative scraping."); Hughes, *supra* note 83 ("[P]rofound insights that could enhance social welfare lie at the intersections of these databases."). The self-reinforcing nature of such market dominance is indicative of so-called "network effects". Spencer Weber Waller, *Antitrust and Social Networking*, 90 N.C. L. REV. 1771, 1787–88 (2012).

145. *But see* hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1118 ("hiQ has presented some evidence supporting its assertion that LinkedIn's decision to revoke hiQ's access to its data was made for the purpose of eliminating hiQ as a competitor in the data analytics field, and thus potentially "violates the policy or spirit" of the Sherman Act."); David Bailey, *The New Frontiers of Article 102 TFEU: Antitrust Imperialism or Judicious Intervention?*, 6 J. ANTITRUST ENFORCEMENT 25, 51 (2018) (""[C]lose link" that may well exist between some (often online) markets on which a firm is dominant, on the one hand, and the gathering of huge amounts of data that is used to reinforce or strengthen the firm's position in the dominated market.").

146. *See* D. Victoria Baranetsky, *Data Journalism and the Law*, TOW CENTER FOR DIGITAL JOURNALISM (Sept. 19, 2018), https://www.cjr.org/tow_center_reports/data-journalism-and-the-law.php [https://perma.cc/6NCS-ZLSJ] ("As data collection becomes increasingly important for investigative journalists in particular, legal experts worry about civil and criminal penalties that exist under the statute [i.e. CFAA]—which has been described by some First Amendment advocates as unconstitutionally vague.").

147. *See* Stucke, *supra* note 133, at 321–22 (The data-opoly can dictate who is granted access to the data and for what purpose, and thereby influence the nature of innovation."); Yadav, *supra* note 125, at 150 ("Investors who must spend money on information infrastructure . . . are likely to be less motivated to spend additional funds on researching information."). In the LinkedIn case, the courts were voicing similar concerns. *See* hiQ Labs, Inc. v. LinkedIn Corp., No. 17-16783, 2019 WL 4251889, at 9 ("If companies like LinkedIn, whose servers hold vast amounts of public data, are permitted selectively to ban only potential competitors from accessing and using that otherwise public data, the result—complete exclusion of the original innovator in aggregating and analyzing the public information—may well be considered unfair competition under California law."); hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1119 ("For present purposes, the Court concludes that the public interest favors hiQ's position. As explained above, the actual privacy interests of LinkedIn users in their public data are at best uncertain. It is likely that those who opt for the public view setting expect their public profile will be subject to searches, date mining, aggregation, and analysis. On the other hand, conferring on private entities such as LinkedIn, the blanket authority to block viewers from accessing information publicly available on its website for any reason, backed by sanctions of the CFAA, could pose an ominous threat to public discourse and the free flow of information promised by the Internet.").

148. *See* David Easley, Maureen O'Hara & Liyan Yang, *Differential Access to Price*

restricting its dissemination in other ways has generally been a precursor for threatening a market equilibrium in which all interested investors have equal incentive to enter the market.[149] Giving, for example, hedge funds – which need not be established players, but may also just consist of a couple of talented and ambitious economists – and their data suppliers the leeway to harvest publicly available data for their endeavors in a way democratizes the digital world.

If the concern is that single data sets – in conjunction with exclusive purchase agreements – become so dominant that whoever purchases them has an unassailable investing edge over every other market participants, antitrust issues might possibly arise and competition law may need to address those concerns if and when they become prevalent. Nonetheless, this does not, in and of itself, make trading on that data fraudulent behavior, it merely raises free market concerns that are not best undertaken by insider trading doctrine.

C. Market Incentives

On the one hand, the use of scraped data by hedge funds seems like an inventive way to make money in the age of seemingly endless amounts of data that ordinary people have difficulty making sense of. Since the data is out there, smart analysts at hedge funds should arguably be able to connect the dots and reap the benefits of their ingenuity.[150] On its face, it seems

---

Information in Financial Markets 51 J. FIN. & QUANT. ANAL. 1071, 1073 (2016) ("If the profit-maximizing price of market data is high, then in equilibrium only some traders purchase price data, and therefore our model exhibits differential access to price information. We show that differential access generally increases the cost of capital and volatility, reduces market efficiency and liquidity, discourages the production of fundamental information, harms liquidity traders, and benefits rational traders relative to an economy in which all traders observe price data simultaneously.").

149. *See* Rory Van Loo, *Rise of the Digital Regulator*, 66 DUKE L.J. 1267, 1285 (2017) ("Barriers to entry arise from the difficulty in obtaining key data. [. . .] Part of the problem is that bots often cannot legally obtain the information they would like to aggregate and analyze, even if that information seems freely available on the web. For online data, sellers have used the law to prevent web scraping, by which intermediaries use web crawlers or spiders to gather online information."). Some even argue that innovation and competition will be helped by regulators stepping in to dilute the IP rights of tech giants. *But see* Dayen, *supra* note 135 ("If Google were put into a compulsory licensing regime, it would have to give up patents for its search algorithms, self-driving cars, mapping software, virtual reality, and Android operating system, to name a few. As the Bell Labs example shows, this type of antitrust enforcement enhances public welfare by benefiting both competition and innovation.").

150. *See* Dirks v. S.E.C., 463 U.S. 646, 659 n.17 (1983) (quote from the SEC briefs: "analysts remain free to obtain from management corporate information for purposes of

entirely counterintuitive to a free market to bar trading on astute analytical findings and thereby put a severe damper on incentives to utilize modern technology in seeking a competitive edge in trading.[151] Whereas investment research used to involve on-the-ground fact-finding and physical investigations in the not-too-distant past, the same – or even better – assessment on a companies' prospects can now be made with the help of digital tools alone.[152] So long as the scraping process does not involve forcibly breaching closed networks or the technologically deceptive circumvention of other technological barriers – since there is a good argument to be made that hacking, as in *Dorozhko*, does pose a threat to an economically desirable level of investor trust[153] –, the proper market incentive should lie with fully exhausting every option at an investor's disposal.[154] With the technological capabilities of investors and their advisors blossoming exponentially, so should insider trading law acknowledge the current state of play in investment research.

From an economic standpoint, such data activism leads to overdue price-corrections in the stock market that further pricing efficiency in the capital markets, thereby helping investors make informed decisions and allocate capital in a pareto-efficient[155] manner.[156] The important part that

"filling in the 'interstices in analysis'"). The law even recognizes a copyright interest in an original selection or arrangement of facts. Feis Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 345, 348 (1991). For a philosophical underpinning of this argument, see Colombatto & Tavormina, *supra* note 55, at 21: "[O]ne can make a Lockean claim in favour of exclusive property rights on the production of knowledge and the use of information (e.g. data treatment and analysis)."

151. Fletcher, *supra* note 127, at 532 ("[I]f a trader had to inform her counterparty of the results of her research before trading, there would be no profits to be earned, nor, more importantly, would there be any incentives to do research.").

152. *See* Salzman, *supra* note 1 (referencing Tim Harrington, CEO of BattleFin, an alternative-data company: "[D]riving to retail phone stores to talk to managers and see what was selling" versus "sell[ing] a product that tracks every time a new phone is turned on and [. . .] find[ing] out quickly 'which wireless carriers are gaining subscribers and which are losing them.'").

153. Karol, *supra* note 66, at 214–17.

154. *See* Stuart P. Green & Matthew B. Kugler, *When is it Wrong to Trade Stocks on the Basis of Non-Public Information? Public Views of the Morality of Insider Trading*, 34 FORDHAM URB. L.J. 445, 484 (2011) (commenting on the outcome of an empirical study carried out by the authors: "In the view of our subjects, merely trading on the basis of an informational advantage not held by other traders does not make such trading blameworthy or deserving of punishment.").

155. *See* Guido Calabresi, *The Pointlessness of Pareto: Carrying Coase Further*, 100 YALE L.J. 1211, 1215–16 (1991) ("[A] society is not at its optimal position if there exists at least one change which would make someone in that society better off and no one in it worse off").

156. *See* Fox et. al., *supra* note 61, at 821–22 ("Informed Trading – trading on information

investment analysts play in bringing about and maintaining efficiency in market pricing through their dissemination of complex and amorphous blocks of information has even been recognized by the SEC.[157] Potentially holding an investor civilly and criminally liable for helping the market discover a mispricing quicker arguably impedes the free and self-correcting nature of the market. Every investor potentially has access to that information, laying to rest the perceived idea that only hedge funds benefit at the expense of every other investor. Barring an economically significant use of scraped data might also impede the existence of various online services, many of which are dependent on users' data being gathered for further use in exchange for free provision of the platform.[158]

In addition, trading on scraped data does not only benefit privileged executives at the expense of ordinary citizens whose privacy rights are allegedly being violated. On the contrary, alternative data is also being used as a source for investment research by pension funds.[159] Hedge funds themselves have pension funds as clients as well, so any significant gains made by the incentivized fund management also benefit future pension prospects for millions of workers.[160] Neither is trading on scraped data

---

or analysis not yet reflected in a stock's price – drives much of the stock market. [. . .] The basics of microstructure economics reveal that informed trading leads to more accurate share prices, which in turn increase the efficiency with which the economy allocates goods and services."). For a far-reaching reform proposal that is based on the idea that free markets are currently impaired by disclosure obligations and insider trading doctrine, see Kevin S. Haeberle & M. Todd Henderson, *A New Market-Based Approach to Securities Law*, 85 U. CHI. L. REV. 1313, 1345–49 (2018). The authors call for an extensive reworking of securities law to allow companies to sell off their soon-to-be-disclosed information to the highest bidder, i.e. the one paying market price, to encourage efficient disclosures as well as a healthy, supply-and-demand-based information flow that deters inexperienced consumers from incurring inefficient trading losses. *See also* Sudipto Bhattacharya & Giovanna Nicodano, *Insider Trading, Investment and Liquidity: A Welfare Analysis*, 56 J. FINANCE 1141, 1155 (2001) ("[I]nsider trading can improve outsiders' welfare").

157. Dirks v. S.E.C., 463 U.S. 646, 659 n.17 (1983) (quoting from the SEC briefs: "[t]he value to the entire market of [analysts'] efforts cannot be gainsaid; market efficiency in pricing is significantly enhanced by [their] initiatives to ferret out and analyze information, and thus the analyst's work redounds to the benefit of all investors.").

158. Rostow, *supra* note 108, at 687 ("[D]ata sale is an enormous, multi-billion-dollar industry that also provides many positive benefits–including the many free services that are offered online").

159. Salzman, *supra* note 1 ("The trend has even been embraced by some pension funds, historically among the most conservative investors"; partially quoting Marcel Prins, COO of APG Asset Management with over $400 billion assets under management: "using such [alternative] data is now 'part of being an active long-term responsible investor'").

160. According to a J.P. Morgan study, around 10% of all assets managed by hedge funds (around $3 trillion in 2018) stem from pension funds. J.P. MORGAN'S CAPITAL ADVISORY GROUP, INSTITUTIONAL INVESTOR SURVEY: INVESTOR TRENDS AND INSIGHTS 4 (2019), https://

limited to behemoth hedge funds: Recognizing current market trends on the basis of smartly accumulated data is, on the contrary, a job perfectly suited for smaller entities as well, lending credibility to the idea that making (legally obtained) data available for trading purposes is an equalizer from a competitive standpoint, not something that deserves to be suppressed.[161] Finally, ordinary consumers might actually benefit from heightened informational asymmetries through the potential reign of data experts as it becomes abundantly clear that they can only lose money against more sophisticated trading competitors and disincentivizes them from actively participating in a market that can, on a statistical average, only work against them.[162]

D. Compatibility with Rationale Behind Insider Trading Doctrine?

What should be the decisive argument is whether the economic rationale behind forbidding insider trading in the first place views trading on scraped data as a scourge to be curbed or as a new-age spectacle to be beheld: If an average investor rightfully feels that the system is somehow stacked against him if trading on scraped data is permitted, he might either refrain

---

www.jpmorgan.com/jpmpdf/1320747018387.pdf [https://perma.cc/Y2W3-CTH9]. APG Asset Management, which manages over $400 billion in assets for Dutch pension funds, invests about 10% of that in hedge funds and private equity. *Top hedge funds at full capacity keeps $548 billion APG away*, BLOOMBERG, June 7, 2018, https://www.pionline.com/article/20180607/ONLINE/180609898/top-hedge-funds-at-full-capacity-keeps-548-billion-apg-away [https://perma.cc/X35T-8NTS].

161. Michael Burry, who — as the Cassandra of the Financial Crisis of 2009 — foresaw the housing bubble based on his own diligent, data-enhanced, research and now manages a small fund, is currently of the opinion that large investors overvalue big firms and, again based on his analytical research, sees significant alpha in becoming stakeholder of small-to medium-sized enterprises, in large part because the market, in his view, lacks ambitious and entrepreneurial small-time investors. Those, through their kinship with the smaller enterprises, might feel more inclined to recognize their oft-hidden value. *See* Heejin Kim & Myungshin Cho, *The Big Short's Michael Burry Sees a Bubble in Passive Investing*, BLOOMBERG, August 28, 2019, https://www.bloomberg.com/news/articles/2019-08-28/the-big-short-s-michael-burry-sees-a-bubble-in-passive-investing?srnd=markets-vp [https://perma.cc/7M5V-ZVTQ] (quoting Burry: "The bubble in passive investing through ETFs and index funds as well as the trend to very large size among asset managers has orphaned smaller value-type securities globally"). All this to say that data scraping may yet enable smaller trading entities, including tech-savvy individuals, to spot structural shortcomings of large funds and other behemoths, thereby correcting a market deficiency that might otherwise have gone unnoticed.

162. Haeberle & Henderson, *supra* note 156, at 1361–68 (calling for extensive securities law reform towards a tiered-access-system to new information on this ground). *See also* Fletcher, *supra* note 127, at 532 ("[T]rading in financial assets is a zero-sum game, the transfer of wealth between parties is expected to be because of the skill or luck of the counterparty").

from participating altogether[163] (thereby contributing to an inefficient allocation of capital as value-generating exchanges on the stock market are foregone in favor of less efficient, but seemingly more transparent trades outside of it) or he might (in most instances, unsuccessfully) try to outsmart the perceived cheaters (again misallocating resources as the investments into gaining unfair informational advantages are inefficient uses of capital).[164] That is not the case here. Investors are merely incentivized to use modern technology for investment research. While some investors might be in awe of other's technological prowess regarding the harnessing of alternative data, that is not to be remedied with insider trading doctrine, but merely signals the competitive edge at the top that capitalism and free markets are built on. This hypothesis is backed up by public opinion: In a recent study, in which an investment analyst was able to profit off of a merger prediction through acquiring "substantial and quasi-exclusive informational advantages" with the help of a sophisticated software tool (evoking now-prevalent scraping trends), the "overwhelming majority" of participants viewed the analyst's conduct as legal.[165] This shows that the public at large can and does differentiate between a competitive advantage based on technological prowess – even if it leads to all but guaranteeing an investment windfall – and one that truly originates in some form of unfair behavior. As anchoring insider trading law in societal perceptions of what should be illegal is a policy directive not to be scoffed at,[166] this further speaks to the legality of data scraping for trading purposes.

Scraped data is by no means fool-proof either: With the market volume of alternative data retail steadily rising, there are also many datasets on offer that offer conflicting or useless information.[167] Even within the alternative

163. *See* Green & Kugler, *supra* note 154, at 454 ("[P]eople are reluctant to 'buy in' to a system that they do not perceive as fair . . . ").

164. For the basis of insider trading liability, see Victor Brudney, *Insiders, Outsiders, and Informational Advantages Under the Federal Securities Laws*, 93 HARV. L.REV. 322, 356 (1979). "A rational buyer (or seller) in a market, who knows that the person with whom he is dealing has material information about the value of the product being exchanged which he could not lawfully acquire, will either refrain from dealing with that transactor or demand a risk premium. If the market is thought to be systematically populated with such transactors some investors will refrain from dealing altogether, and others will incur costs to avoid dealing with such transactors or corruptly to overcome their unerodable informational advantages." *Id*.

165. Green & Kugler, *supra* note 154, at 480–81.

166. For further details on the desirable link between social norms and the threat of legal sanctions, see Green & Kugler, *supra* note 154, at 450.

167. *See* Dumont, *supra* note 36, at 18 ("[C]ombining these less-vetted sources with processing systems that few understand can also downplay truly material information and focus too much attention on the noise"); Egan, *supra* note 3 (quoting Dev Kantesaria, founder

data market, intelligence, industry and sector-specific knowledge and independent research still remain key to correctly assess any scraped data aggregation.[168] That means the typical mechanism of insider trading – one party receives or misappropriates information in violation of a fiduciary duty that all but guarantees illicit gains on the stock market[169] – is not even in (full) force here.

E. Interim Conclusion

As described above, accessing public information (even with technological means that are not appreciated by the host of the information) and indexing that data in a way that delineates a point that was previously drowned out by the sheer abundance of information simply does not constitute fraud on the market. Rather, it represents the product of one of the most important skills in the twenty-first century going forward: The ability to disseminate large portions of information in a way that cuts through all the noise and illuminates a meaningful point while doing so. That is something that should be encouraged rather than disincentivized. Trying to preserve an unnatural informational equality of market participants by wielding the heavy hammer of Section 10(b) and Rule 10b-5 and practically

---

of Valley Forge Capital Management: "I find the usefulness of datamining techniques to be a crapshoot"); Horton, *supra* note 11, at 9 (quoting Evan Schnidman, CEO of Prattle: "There is a lot of data out there that has been under-analyzed or incorrectly analyzed due to spurious correlations"); INTEGRITY RESEARCH ASSOCS., *supra* note 39, at 16 ("Alternative data may be scarce or hard to obtain but that does not necessarily mean it will contribute positively to the investment process."); Ian King, *Hedge Funds Find an Edge with Big Data*, BANYAN HILL, December 7, 2018, https://banyanhill.com/hedge-funds-big-data/ [https://perma.cc/D2ZF-ZPGL] ("[A]s the industry gains popularity, and everyone starts using the same data, its impact on investment returns will be muted"); Kochkodin, *supra* note 42 (quoting George Mussalli, CIO at PanAgora Asset Management: "We stay away from over-marketed data purely curated for hedge fund consumption, such as satellite data, credit card transactions, and email receipts. These data sources are overused, and we have seen a marked deterioration in their-predictive power."); Rostow, *supra* note 108, at 683 ("Not all information is useful."); Wieczner, *supra* note 3 ("One large quant hedge fund got stung when its algorithm confused sarcastic tweets about Lululemon's (LULU) see-through pants debacle with positive sentiment, buying shares in the yoga-apparel retailer when it should have been selling."). *See also* Huan Liu, Fred Morstatter, Jilian Tang & Reze Zafarani, *The Good, The Bad, and The Ugly: Uncovering Novel Research Opportunities in Social Media Mining*, 1 INT. J. DATA SCI. ANAL. 137, 137–38 (2016) (citing examples of problematic data, such as " . . . bias in social media, data, evaluation dilemma, data reduction, inferring invisible information, and big-data paradox.").

168. Kochkodin, *supra* note 42 (quoting Ray Iwanowski, Principal at Secor Asset Management LP: "[a]vailable information is not synonymous with useful information").

169. For a symbolic example, *see* the wording in *United States v. Martoma*, 894 F.3d 64, 76 (2d Cir. 2017) ("[Stating] 'You can make a lot of money by trading on this' is strong circumstantial evidence of the tipper's intention to benefit the tippee").

outlawing analytical entrepreneurism in the process is tantamount to willfully dumbing down investors. That cannot be in the best interest of a free market.

V. CONCLUSION

Are investments made on the basis of scraped data at risk of being considered insider trading? After review of the judicial standards and some extrapolation on questions without on-point case law, the answer is no, as long as the information aggregated within those data sets is not the product of hacking (i.e. breaching authentication barriers through technological means of deception) or does not contain information that can only be obtained through fiduciary duty violations. Therefore, the only thing that investors should be very careful about, is aggregating or purchasing data sets from third parties that contain information about companies that can't possibly be available in the public sphere (e.g. proprietary customer data) or obtain or purchase information that has been accessed through aggressively dubious technological means.

Might trading on scraped data be illegal sometime in the near future? As with every difficult prognostic question that involves some reading of (legal, in this case) tea leaves, the answer to this policy question is: It depends – and by no means least on one's views on the best way for (free) markets for operate. This article posits the view that any policy concerns with data scraping should be handled by the pertinent substantive law (e.g. data protection law to combat privacy issues, antitrust enforcement actions to combat competition concerns etc.) rather than abusing insider trading law for exactly that. Furthermore, bringing down the hammer of insider trading doctrine on investments made in reliance on scraped data would fossilize the current hegemony of digital behemoths and disincentivize analytical data entrepreneurism, both severe impairments to the free market the U.S. aspires to be. The only certain thing is that technological evolution is currently revolutionizing investment research – and insider trading doctrine should quickly adapt to these changes to avoid being either over-inclusive or becoming entirely irrelevant.