

Accepted for publication in *Theory, Culture & Society*

How Algorithms Interact: Goffman's 'Interaction Order' in Automated Trading

Donald MacKenzie

June 2018

Author's Address:
School of Social and Political Science
University of Edinburgh
Chrystal Macmillan Building
Edinburgh
EH8 9LD
Scotland
DonaldMacKenziePA@ed.ac.uk

How Algorithms Interact: Goffman's 'Interaction Order' in Automated Trading

Abstract

In a talk in 2013, Karin Knorr Cetina referred to 'the interaction order of algorithms', a phrase that implicitly invokes Erving Goffman's 'interaction order'. This paper explores the application of the latter notion to the interaction of automated-trading algorithms, viewing algorithms as material entities (programs running on physical machines) and conceiving of the interaction order of algorithms as the ensemble of their effects on each other. The paper identifies the main way in which trading algorithms interact (via electronic 'order books', which algorithms both 'observe' and populate) and focuses on two particularly Goffmanesque aspects of algorithmic interaction: queuing and 'spoofing', or deliberate deception. Following Goffman's injunction not to ignore the influence on interaction of matters external to it, the paper examines some prominent such matters. Empirically, the paper draws on documentary analysis and 185 interviews conducted by the author with high-frequency traders and others involved in automated trading.

Key Words

Interaction order; Erving Goffman; Karin Knorr Cetina; algorithm; high-frequency trading; queuing; spoofing.

‘[H]uman awareness comprises the tip of a huge pyramid of data flows, most of which occur between machines’ (Hayles, 2006: 165).

As Hayles points out, human beings are increasingly enmeshed in a ‘cognisphere’, shared with machines, in which many important processes take place among those machines, without direct human involvement. How should what Beer (2009: 987) calls ‘the technological challenges to human agency offered by the decision-making powers of established and emergent software algorithms’ be theorised? This paper addresses this question for one specific area: automated financial trading, especially high-frequency trading or HFT, which is ultrafast and involves very large numbers of trades.

The paper takes up a suggestion made by Karin Knorr Cetina in a talk to the panel ‘Theorizing Numbers’ at the American Sociological Association, in which she used the evocative phrase: ‘the interaction order of algorithms’ (Knorr Cetina, 2013). It points us in a somewhat different direction to much recent work on algorithms, which draws upon theorists as sophisticated and well-known as Hayles herself (e.g. 1999, 2012; see also Gane, Venn and Hand, 2007), Foucault (e.g. Cheney-Lippold, 2011 and Bucher, 2012), Deleuze (e.g. 1992: see e.g., Savat and Poster, 2005 and Cheney-Lippold, 2011), Latour (e.g. 2005) and Lash (2002, 2007; see Beer, 2009).

The term, ‘the interaction order’, was coined by Erving Goffman, whose primary reputation is not as a theorist – even a critic as sympathetic as Burns (1992) could find his theorising unsystematic and sometimes even careless – but as a hugely insightful observer of social interaction. ‘The Interaction Order’ was the title of Goffman’s intended 1982 Presidential Address to the American Sociological Association, undelivered because he was already suffering from the cancer that was soon to kill him, but published the following year (Goffman, 1983). In it, he laid out what he saw as most central to his life’s work:

Social interaction can be identified narrowly as that which uniquely transpires in social situations, that is, environments in which two or more individuals are physically in one another's response presence. (Presumably the telephone and the mails provide reduced versions of the primordial real thing.) ... My concern over the years has been to promote acceptance of this face-to-face domain as an analytically viable one – a domain which might be titled, for want of any happy name, the *interaction order* (Goffman, 1983: p. 2, emphasis in original).

The uneasy parenthesis in that quotation points to the need to question the primary role of physical co-presence in Goffman's conception of the interaction order. In the decades since 1983, 'the telephone and the mails' have been joined by multiple other forms of mediated communication: electronic mail, text messages and other forms of instant messaging, social media, Skype and other forms of telepresence, etc. As these have grown in importance, Knorr Cetina is surely right to suggest supplementing Goffman's focus on spatial proximity with a broader, temporal notion of 'response presence' as accountability 'for responding without inappropriate delay to an incoming attention or interaction request' (Knorr Cetina, 2009: 74).

Given this paper's focus on algorithmic trading, it is particularly relevant that both Knorr Cetina herself and Alex Preda have productively deployed reworked versions of Goffman's 'interaction order' to analyse human beings trading electronically. Much of Knorr Cetina's research on financial markets has concerned foreign-exchange dealers in bank trading rooms communicating with other traders (in different banks, but personally identifiable and sometimes personally known) via the Reuters 'conversational dealing' system, an early electronic system – still in use today – that combines automated requests for price quotations with the capacity to formulate Telex-style messages conveying up-to-date market information, pleasantries ('please', 'thanks'), and the details needed to settle trades (see, especially, Knorr Cetina and

Bruegger, 2002a). However, Knorr Cetina also examines human traders interacting with a fully anonymous electronic market (e.g. Knorr Cetina, 2009: 72-73), as does Preda (2009 and 2013). In the work of Knorr Cetina and Preda, Goffman's notion of the interaction order gets stretched beyond temporal response presence among spatially separate but identifiable humans, as 'the market' itself becomes a party to 'postsocial' interaction (Knorr Cetina and Bruegger, 2002b). As Knorr Cetina points out, in projecting 'the market', traders' computer screens project 'an "other" for participants, with whom these participants interact' (Knorr Cetina, 2009: 73; see also Knorr Cetina and Preda, 2007). Preda discovers human traders – no longer in trading rooms, but often physically entirely alone – trying to disaggregate 'the market' into different kinds of agent (for example, 'an individual [human] trader, an institution, or a robot': i.e., a trading algorithm) that do different things, and sometimes (even though alone) audibly addressing these absent, imagined, unhearing others, 'engaging with "guys", "dudes", and "buds"' (Preda, 2013: 42; Preda, 2009: 687).

Knorr Cetina's invocation of 'the interaction order of algorithms' invites us to take yet a further step, which is this paper's focus: to extend the notion of 'interaction order' to situations in which trading algorithms interact with each other rather than with human beings. First, though, we need to be clear what 'algorithm' means in this context, and what it might mean for algorithms to interact. I follow how my interviewees use the term 'algorithm'. For them, algorithms are not simply the abstract 'effective procedures' (finite sets of exact, 'mechanical' instructions) of metamathematics or computer science. Rather, an 'algorithm' is a material implementation of such a procedure: i.e., a computer program running on a physical machine.

Although this view of algorithms is implicit in much of the literature pointed to above – for example, in Lash's discussion of '[p]ower through the algorithm' (2007: 71)

– it is worth spelling out explicitly that an algorithm is a material entity that does things materially: ultimately, electrically. (The need for speed in automated trading means that there is a sense in which those involved in it *have* to be materialists. For example, they cannot successfully conceive of computers as abstract machines, but have to think of them as assemblages of metal, plastic and silicon through which electrical signals pass: see MacKenzie [2014a]. This points to the relevance here of theoretical traditions in which materiality is prominent, such as ‘media materialism’ [e.g., Kittler, 2006; Parikka, 2015].) Among the things an algorithm does, in automated trading, is to have material effects on the behaviour of other algorithms; reciprocally, their behaviour influences what it does. The ensemble of such effects is what I mean by the ‘interaction order of algorithms’.

Goffman was a thorough-going, albeit tacit, materialist. Human bodies, their positioning, their physical settings, their gestures, glances, blushes, etc, are prominent in his work: see, e.g., Goffman (1959, 1963, 1967 and 1968). The reader’s intuitions may, however, rebel against the application of Goffman’s ‘interaction order’ to the mutual effects of algorithms. Their ‘silicon bodies’ differ radically from human flesh, and they interact explicitly and instrumentally, not subtly and expressively as humans do. And, of course, as far as we know, trading algorithms have no self-consciousness, while humans are often painfully self-aware.

Intuitions nevertheless need to be interrogated. The success with which Knorr Cetina and Preda have applied their extended conceptualisations of the ‘interaction order’ to human beings trading electronically and anonymously suggests that we should not reject *a priori* the notion’s application to trading by algorithms. After all, the information and forms of action available to human beings in most of today’s anonymous electronic markets are often no different from those available to algorithms. Both humans and algorithms face much the same tasks (especially the task of drawing

inferences from the 'order books' described in this paper's second section) and they act in the same way, by entering, cancelling, or sometimes modifying orders, even if they do it with different tools: humans using visual interfaces, keyboard and mouse; algorithms employing direct, computer-to-computer communication.

This paper therefore asks the reader to suspend intuitive judgement while it follows Knorr Cetina's pointer and experiments with applying Goffman's 'interaction order' to automated trading. The empirical material drawn on is research by the author on automated trading (especially on high-frequency trading, but also, for example, on the 'execution algorithms' used by institutional investors to split up big orders), on the exchanges and other trading venues on which it takes place, and on its technological underpinnings. In total, 185 interviews have been conducted, mainly in Chicago and New York, with the developers of trading algorithms, the traders who use them (who are often the same people), exchange staff, providers of technological services, regulators, etc. These interviews (which covered both the current practices of automated trading but also – when the interviewee had had a long enough career to have first-hand experience of this – the historical processes that have shaped current practices) have been supplemented by participant observation at two industry meetings, a visit to Cermak (a data centre in Chicago that houses much algorithmic trading), and examination of web-based discussion forums, of the technical literature, of trade press, of enforcement actions by regulators, etc.

Five sections follow this introduction. The first sets the stage by drawing on this empirical research to describe the physical settings within which trading algorithms interact and to identify the most important way in which they do so. Next comes a section on a form of interaction discussed in Goffman's Presidential Address (and also prominent in ethnomethodological analyses such as Livingstone, 1987) that is of huge importance in automated trading: queuing. Then follows a discussion of one of

Goffman's most persistent concerns: dissimulation, including a form of it particularly salient for automated trading, 'spoofing'. That section includes a discussion of a fascinating episode in which algorithmic action at odds with 'normal' behaviour in queues has formed the basis of an accusation of spoofing. The paper's penultimate section takes up Goffman's reminder not to neglect 'the dependency of interactional activity on matters outside the interaction' (Goffman, 1983: 12) by examining some of the most important of those matters as they bear upon algorithmic trading. The paper's conclusion is, I hope, appropriately modest: it argues that Goffman's 'interaction order' points us in the right direction when studying trading algorithms, but it also identifies the methodological difficulty of research on how trading algorithms interact.

How Trading Algorithms Interact

As already emphasised, this paper views trading algorithms materially, as programs running on trading firms' computer servers. Many, perhaps most, of those servers are to be found in no more than fifteen computer data centres worldwide, in each of which thousands of trading algorithms may be running at any one time. Some of these centres are owned by exchanges such as the New York Stock Exchange; others are multi-user buildings, such as Chicago's Cermak, NY4 in Secaucus, New Jersey, and LD4 in Slough. Cermak used to be a giant printworks (the Sears Roebuck catalogue was printed there: see MacKenzie, 2014b), but most other trading datacentres are purpose-built, and easy to mistake for warehouses. They contain few human beings, mainly security and maintenance personnel. Huge amounts of energy flow into datacentres in the form of electricity, and flow out as heat extracted by powerful cooling systems (tens of thousands of computer servers packed close together generate a lot of heat). Those servers are housed on racks in rows of cages: normally wire-mesh, but sometimes with opaque doors for privacy. Above the cages is a giant spider's web of copper and fibre-optic cables that connects servers to each other (and carries fibre-optic, microwave and

satellite signals from the outside world). Some of the cages contain the servers and switches that make up the computer systems of exchanges and other organised trading venues; other cages contain the servers of the firms trading on those exchanges. The reason for the clustering into a remarkably small number of very big buildings is trading firms' desire to have their servers 'co-located': placed as close as possible to exchanges' systems.

With limited exceptions, the trading algorithms running on these servers do not interact directly with each other, but indirectly, most commonly via an exchange's computer system, and in particular via an electronic file called the exchange's 'central limit-order book', or more simply, its 'order book'. (To avoid cluttering the text, I have gathered together the main exceptions to its empirical generalisations in Appendix 1.) A pictorial representation of a typical – but hypothetical, because I want to use it to illustrate a variety of points as clearly as possible – order book is in Figure 1. It is an order book for shares, but (with exceptions briefly described in Appendix 1) the trading of futures, foreign exchange, U.S. Treasury bonds and stock options is similar in form. On the left-hand side of Figure 1 are the bids to buy the shares in question: for example, there is a bid to buy 100 shares at \$44.99; a bid to buy 44 shares, also at \$44.99, etc. On the right-hand side are the offers to sell, for example an offer to sell 100 shares at \$45.00.

No human traders are to be found in datacentres such as Cermak: humans are in that sense on the periphery of today's trading. A trading algorithm that is housed in a datacentre enters bids or offers into the order book (or cancels, or sometimes modifies, bids or offers it has previously entered) by instructing the network interface card of the computer server on which it is running to send an electronic message through the cable – typically of the order of 100 metres long – that threads its way through the spider's web and connects the server to the exchange's computer system. That system contains

programs called 'matching engines', which process these incoming messages and update the order books for the shares or other financial instruments being traded. If a matching engine finds a 'match' (a bid to buy a financial instrument, and an offer to sell it, both at the same price) it executes a trade; otherwise, it simply adds new bids and offers to the order book.

As well as trading algorithms sending in the bids and offers that populate the order book, they also 'observe' it (my term, not interviewees'). Whenever a matching engine receives a new order or a cancellation or modification of an existing order, or it finds a match, it sends the exchange's feed server a message containing the anonymised details. That server then disseminates these messages to subscribers to the exchange's datafeed. (The 'hidden orders' mentioned in Appendix 2 are, however, not disseminated.) The datafeed flows – again through around 100 metres of fibre-optic cable – to trading firms' servers, which use the stream of messages to construct their own electronic 'mirrors' of the order book.

Trading algorithms interrogate this mirrored order book in a variety of ways, seeking to predict price changes. In the order book in Figure 1, for example, there are offers to sell 4,240 shares, and bids to buy 1,324; 'supply' thus exceeds 'demand', and thus a fall in price might be predicted. While no sophisticated trading algorithm would rely on a calculation as simplistic as this, interviewees reported heavy reliance by algorithms on various forms of weighted average of the numbers of financial instruments being bid for and offered at different prices, along with a variety of ways of inferring the dynamics of how the order book is changing through time. The pervasive concern, discussed below, with 'spoofing' means that sophisticated trading algorithms will also deploy various means of assessing the likelihood that the existing bids and offers in the order book will actually be cancelled before they are executed, and will discount those for which this is the case.

Predictions based on these algorithmic ‘observations’ of the order book (along with similar observations of the order books for other instruments whose prices are known to be correlated with those of the instrument being traded) are used for two main forms of profit-seeking trading. The conceptually simpler is ‘liquidity-taking’ or ‘aggressive’ trading. Suppose an algorithm’s observations generate the inference that the price of the shares being traded via the order book in Figure 1 is about to fall. It could then send to the matching engine an order to sell shares at \$44.99, which the matching engine can execute at least in part as soon as it has processed it, because it can match it with existing bids to buy at \$44.99. (That is why it would be called a ‘liquidity-taking’ order: it removes an existing order or orders from the order book.) If the price does indeed fall below \$44.99, then the algorithm can buy back shares at a profit.

‘Liquidity providing’, in contrast, involves an algorithm sending the matching engine orders that cannot immediately be executed, and its most systematic form (known as ‘market-making’) involves continually keeping both a bid and a higher-priced offer in the order book, in the hope that both will be executed and the difference in their prices captured as profit. Suppose, for example, that in Figure 1 the same algorithm has entered into the order book both the bid to buy 100 shares at \$44.99 and the offer to sell 100 shares at \$45.00. If both are executed, the algorithm will make a profit of one cent for each share traded. That sounds negligible, but high-frequency trading involves the buying and selling of huge numbers of shares, so tiny profits add up.

A market-making algorithm has just as much need as a liquidity-taking algorithm electronically to ‘observe’ the contents of the order book and thus to predict price movements, because if prices move sharply it can easily be left with an inventory of shares the prices of which have fallen, or with what participants call a ‘short position’ in shares whose prices have risen.¹ The constant observation of the order book by

¹ That is to say, it may have sold shares that its firm does not own.

trading algorithms of all kinds, and the actions they frequently take in response to that observation, mean that an explicit 'global' electronic representation – a representation of the entirety of 'the market' in question – plays a much larger role than in most ordinary human social interaction. (At a party, for example, most participants' attention is devoted to a small subset of what is going on, with only an anxious host or hostess maybe monitoring the event as a whole: see, e.g., Goffman 1963.)

As Yuval Millo pointed out to me in a personal communication, the crucial role of a global representation in algorithmic trading suggests the need for nuance, when analysing it, in invoking metaphors – such as 'swarms' (see Vehlken, 2013) – in which there is self-organisation resulting from local interactions, for example between nearest neighbours. (There *are* some local interactions among trading algorithms: see Appendix 1.) Again, though, the central role of a global representation is fully consistent with Knorr Cetina's and Preda's extensions of Goffman's 'interaction order'. The human traders they studied also devote much or sometimes even all of their attention to a global representation on screen of the overall market, a representation that today is usually simply a computer file presented in a form (such as Figure 1) suited to human eyes. Like algorithms, those human traders also simultaneously observe and construct the object of their attention.

Queuing

After sketching overall features of the human interaction order, Goffman (1983: 6) went on 'to try to identify the basic substantive units, the recurrent structures and their attendant processes' asking '[w]hat sort of animals are to be found in the interactional zoo?' Among his first examples was the queue: '[w]hat queues protect is ordinal position determined "locally" by first come first placed' (Goffman 1983: 16).

That ordering is precisely the one enforced by most matching engines (for the main exceptions, see Appendix 1). For example, the offer to sell 50 shares at \$45.00 in

Figure 1 will be executed only once the earlier offer to sell 100 is executed or cancelled. It is natural to conceptualise this ordering as a 'queue', and that is how participants do indeed think of it. Queues are of huge importance in automated trading; Pardo-Guerra (forthcoming) summarises the field's history as 'from [trading-floor] crowds to queues'. As those of my interviewees who had traded manually in Chicago's trading pits reported, an ordering similar to the start of a queue did often emerge in those crowds. In a pit, bids and offers were either shouted out or hand-signalled, and were thus observable to the traders crowded into the pit. While a variety of factors – including informal 'sharing' norms and reciprocity – affected who got which trade, there was often agreement as to which trader had, for example, made the first bid at a given price, and an informal convention that s/he then deserved to have that bid executed first. This limited form of ordering was, in classically ethnomethodological fashion, 'reflexive, self-organizing, organized entirely *in situ*, locally' (Livingston, 1987: 10). In automated trading, however, queues are not simply self-organised: they are structured electronically by exchanges' matching engines.

Queue position is not a pressing concern for 'aggressive' algorithms (liquidity-taking orders don't usually encounter queues), but it matters enormously to market-making algorithms' liquidity-providing orders. If these orders are too far back in the queue, they may simply never be executed, and so no profit will ever be made. Getting to the front of the queue is a matter of technical expertise (such as the 'close-to-the-metal' programming, as participants call it, needed to speed processing by a computer as a physical machine) and of spatial location. Queue position is one chief reason why trading firms pay exchanges to co-locate their servers alongside the exchange's computer system. Speed, and therefore queue position, can, however, also be achieved more informally. Before the electronic messages containing orders reach the matching engine, they are processed by order gateways. These are normally identical computer servers, running identical software, and identically linked to the matching engine.

However, each gateway typically serves more than one trading firm, and if a firm has to share a gateway with a firm whose algorithms send in large numbers of orders, the former's algorithms' orders may be delayed. Avoiding this can be a major practical issue; it is, for example, helpful (I was told by a former high-frequency trader) to know exactly whom to speak to at the exchange should it happen. 'If you didn't know to call that person, you'll start at some low-level help-centre desk'.

There are also other subtleties to algorithmic queueing, which go beyond the need for speed, and which are sometimes deeply controversial among insiders to the world of automated trading. As both Goffman and ethnomethodologists such as Livingston (1987) emphasised, the interaction order of human queues is a moral order: first come first served 'produces a temporal ordering that totally blocks the influence of such differential social statuses and relationships as the candidates bring with them to the service situation' (Goffman, 1983: 14). Especially in US share trading, a variety of types of bids and offers are available to some algorithms (but not always to others), which can be used to help an algorithm get to the front of the queue: see Appendix 2. These bids and offers have generated much controversy (both among my interviewees and also in public forums: see, e.g Bodek 2013). The accusation against them has in effect been that they allow 'differential social statuses and relationships' illegitimately to influence queue position.

Dissimulation

As noted, one of Goffman's persistent interests was the role of dissimulation in interaction. He was, of course, no naive moralist, and fully understood that presenting a false impression is sometimes entirely appropriate (it is, for instance, right for a medical student who is nervous to hide that fact when treating a patient) and that 'tact' – for instance, pretending not to notice an occurrence that would cause a participant to lose 'face' – is often desirable.

Algorithms, too, dissimulate. Consider the excess of offers to sell in the order book in Figure 1. Much of it is made up of three large offers (for 1000, 400 and 700 shares) with prices that are at least two 'levels' away from the best offer price of \$45.00. Under normal circumstances, the algorithm (or, perhaps, even human being) that has posted those offers will have the time to cancel them before they are executed. So maybe they have been entered into the order book so as to produce an excess of offers relative to bids, and thus cause other algorithms to predict a price fall and therefore to sell. The original algorithm can then profit from the price decline it has caused, for example by buying at a temporarily low price, cancelling the large offers, and selling when prices recover.

For an algorithm or human to do that is what market participants call 'spoofing'. It is, for example, what the west London trader, Navinder Singh Sarao, who was arrested in April 2015, is accused of by the US Department of Justice. Its indictment quotes emails allegedly sent by Mr Sarao in which he requested technical help adding a particular feature to his trading software, 'a cancel if close function, so that an order is canceled if the market gets close', with a further refinement to permit him 'to be able to alternate the closeness ie one price away or three prices away etc etc' (US Department of Justice, 2015: 7-8; in Figure 1, an offer to sell at \$45.01 is 'one price away' from the best offer).

Given that spoofing is illegitimate and generally now illegal (see below), it is unsurprising that none of my interviewees admitted to writing algorithms that spoofed. They did, however, talk about how important it was for any algorithm that made price predictions on the basis of an analysis of the order book to be able to distinguish 'real' orders in that book from 'spoofer' orders that would be cancelled before being executed. One of them had, for example, programmed his firm's algorithms to give less weight to a single big order than to multiple small orders of the same aggregate size, because the former was less likely to be 'real'. Both he and another interviewee were experimenting

with artificial-intelligence machine learning techniques – especially ‘support vector machine’ techniques – to make their algorithms more sophisticated in how they distinguished ‘real’ from ‘spoofer’ orders. (One of the surprises of the interviews with the designers of high-frequency trading algorithms is the otherwise rather limited use of artificial-intelligence techniques in price prediction. HFT algorithms, especially market-making algorithms that have to get to the heads of queues, often employ conceptually very simple but ultrafast inferences, such as ‘weighted’ counts of bids and offers or extrapolation to the stock market of movements in the market for stock-index futures. Liquidity-taking algorithms, which can afford to act a little more slowly, do employ more sophisticated inferences, but interviewees at firms that specialised in these algorithms reported that the patterns in order-book dynamics they exploited were often at the border of statistical significance, and the low signal:noise ratio caused difficulties for machine-learning techniques.)

What is, from the viewpoint of this paper, a particularly interesting set of instances of alleged spoofing was described to me by an interviewee in June 2015. In all the previous examples of spoofing I had encountered, the alleged ‘fake’ orders were placed not at the best bid or offer, but one or more levels away from it. The new set concerns orders *at* the best bid or offer price, such as the offer to sell 600 shares at \$45.00 in Figure 1.

For an algorithm to place a fake order at the best bid or offer price is potentially an effective means of moving a market, because algorithms that make inferences based on counts of the contents of the order book typically (so interviewees told me) ‘weight’ these orders more heavily than orders further away, partly because those latter orders have traditionally been more likely to be fake. (An algorithm summing the offers in Figure 1 might assign a weight of 1.0 to the offers at \$45.00; a weight of 0.5 to offers at \$45.01; 0.25 to offers at \$45.02; etc.) However, a fake order at the best bid or offer price

is also dangerous to the intended spoofer, because it is much more likely to be executed before it is cancelled (it would be particularly dangerous for a slow human being rather than a fast algorithm to attempt to spoof in this fashion).

What first led my interviewee's firm to suspect spoofing was behaviour at odds with the normal interaction order of queuing. It involved apparent use of the 'modify up' instruction in the electronic trading system of the exchange in question. That instruction alters an existing bid or offer by increasing the number of financial instruments being bid for or offered. If this instruction is employed, the order that has been modified goes the back of the queue (as in the case of the offer of 600 shares at \$45.00 in Figure 1). 'You should never do that in a FIFO market', said my interviewee. (FIFO is the acronym of 'first in, first out', and refers to the form of queuing discussed in this paper, in which the first order at a given price received by the matching engine is executed first.)² Doing something that caused an order to lose queue position 'looked weird to us', the interviewee reported. One interpretation might have been that this was 'incompetent' or 'maladjusted' (Livingston, 1987: 14) queuing behaviour, but my interviewee's firm took it to be evidence of spoofing. By using 'modify up', if necessary repeatedly, an order could be kept at the back of the queue, which is of course exactly what an algorithm that is spoofing needs to do to reduce the risk of the order being executed.

Fascinating as spoofing is, it does not exhaust the possibilities of algorithmic dissimulation. Execution algorithms are, as noted above, used by institutional investors to split up large orders; along with high-frequency trading, they are the other most important form of algorithmic trading. Their entire rationale is as a form of dissimulation: the goal is for as long as possible to hide the fact that a big 'parent' order (perhaps for a million or more shares) is being executed, by splitting it into 'child'

² His implicit contrast is with the 'pro rata' markets mentioned in Appendix 1, in which 'modify up' can be employed without detrimental effects.

orders for as few as 100 shares. As an interviewee who headed a major enterprise providing execution algorithms put it:

we'll take that huge order and chop it up into little tiny pieces, and if we do it right anyone who's looking at it can't tell that there is a big buyer: it looks like tiny little retailish trades [the sort of trades a lay investor might engage in] ... My job is trying to obscure what my institutional clients are trying to do, you know, so our role in the market place is to make it so no-one can work out what the hell's going on.

Unlike spoofing, this form of dissimulation is not merely legal but viewed as entirely legitimate. Indeed, the most common form of moral framing in debate over high-frequency trading (see, e.g., Lewis 2014) is to distinguish the 'good' algorithms and technical systems that hide big orders from the 'bad' HFT algorithms that seek to detect the big parent order and change their pricing and order submission behaviour appropriately. That framing, however, is contingent and contestible. Thus, one interviewee, exasperated with what he took to be its facile moralism, reversed it: 'I don't think the guy who's trying to hide the supply-demand imbalance [by employing an execution algorithm], why is he any better of a human being than the person trying to discover what the true supply-de[mand imbalance is]?'

['T]he dependency of interactional activity on matters outside the interaction'

For Goffman, interactions have their own logics and processes, and interaction is 'a particular kind of activity', which is what warrants speaking of 'the interaction order' just as one might refer to 'the economic order' (Goffman, 1983: 5). Goffman, however, also rejected what he called 'a rampant situationalism' (1983: 4). He emphasised repeatedly that, in words already quoted above, what goes on in interaction depends 'on matters outside the interaction', including social relationships and social structure. Although his discussion of how situations and structures interrelate is not as fully

developed as one might wish (see Burns, 1992), the broad outlines of Goffman's account are clear. There is only a 'loose-coupling' relationship (Goffman, 1983: 12) between situations and social structure, but the latter is a real phenomenon, not reducible to an aggregate of multiple interactions. Social relationships and social structure *shape* interactions, but not deterministically: for example, the theoretical interest for Goffman of the queue is (as indicated above) precisely that it is a form of interaction in which their influence is, locally, blocked.

Let me, therefore, follow Goffman and give three examples of the 'loose-coupling' shaping of algorithmic interaction by 'matters outside' it. The first is the changing status of spoofing. When I began interviewing in 2010, spoofing seemed a routine market practice, at least in futures trading: 'most new orders [in the futures market] are fake', a trader in Chicago told me in 2014. There was a long tradition of spoofing being acceptable – in Chicago's trading pits, I was told by another interviewee, a successful spoofer was even admired, much as a skilled bluffer in poker would be – and a tolerant attitude continued in the early years of the transition to electronic trading (Zaloom, 2006; Arnoldi, 2015). Recently, however, disapproval has grown sharply, even though two of the more libertarian-minded of my interviewees still felt strongly that it was quite wrong for the state to try to take action against spoofing. Until 2014, traders who had engaged in spoofing had only ever been subject to administrative action, and the resultant fines could in effect be considered a business expense. However, the Dodd-Frank Act (the main post-crisis legislation in the US) weakened the legal tests that have to be passed for a criminal prosecution for spoofing to succeed, and in October 2014 the first such prosecution began. The trader who told me about the extent of fake orders also reported that in the three weeks since the indictment, the incidence of spoofing, as detected by his firm's algorithms, had gone down sharply.

The second example concerns the shaping of queueing in US share trading by Federal regulation. As summarised in Appendix 2, US stock exchanges are not free to have their matching engines structure queues as they wish. Instead, matching-engine behaviour is governed by Regulation NMS [National Market System], which, although first implemented only in 2007, has roots that can be traced back to the late 1970s (Pardo-Guerra forthcoming). Back then, the Securities and Exchange Commission – long suspicious of the dominance of one exchange, the New York Stock Exchange (NYSE) – sought, with a mandate from Congress, to create a National Market System that would promote competition without leading to market fragmentation. Two designs for that system contended. One, backed by prominent economists, was for a single, national electronic order book to which all brokers and exchanges would send their orders. Unsurprisingly, the NYSE and most of the more minor exchanges saw this proposal as a threat to their existence, and successfully promoted an alternative model in which they would continue to operate much as they did, but linked by a computer network that could be built quickly and easily using existing NYSE technology. Forty years on, that remains the basic structure of US share trading. The different exchanges are still not fused into a single order book. Instead, Regulation NMS's elaborate rules are still seen as necessary to competition.

It is difficult to read this history without thinking of the prescient analysis of neoliberalism in Foucault's lectures on 'The Birth of Biopolitics', delivered (as it happens) in 1979, just as the crucial decisions were being taken as to how to create more 'competition' in US share trading. Competition is not a natural condition, the Ordoliberalists believed: rather, it has to be 'produced by an active governmentality' (Foucault, 2008: 121). Although the influences on it have been more diffuse, the Securities and Exchange Commission has been the chief vehicle of that governmentality in US financial markets, and by constraining how matching engines organise queues it has significantly shaped the interaction order of algorithms.

My third example is a domain of automated trading in which there has been no analogue of that project of governmentality: foreign exchange. (Financial regulations are still largely primarily national in scope, while foreign exchange is intrinsically an international activity that therefore falls into a gap in regulatory coverage.) In foreign exchange, the traditionally dominant actors – the big global commercial banks – have retained, at least until very recently, a degree of market power that banks have largely lost in other exchange-based trading. However, weighed down by old ‘legacy’ software systems, and frequently bureaucratic, big banks are often not good at the development of the fast, sophisticated algorithms needed for HFT. When high-frequency trading of foreign exchange began, the algorithms deployed by small HFT firms therefore found plentiful opportunities for profitable aggressive trading, often at the expense of banks’ slower systems. Banks, however, were able to exert influence on trading venues that had the effect of shutting off many of those opportunities and thus rendering liquidity-taking unprofitable. They have, for example, demanded (often successfully) that their market-making algorithms be granted ‘last look’ privileges: in other words, matching engines grant their algorithms a few hundredths a second – a tiny period for humans, but an eternity for HFT’s fast machines – in which to decide whether to permit the matching engine to consummate a trade. Last look and other measures to constrain liquidity-taking by HFT algorithms have shifted the ecology of algorithms in foreign exchange: interviewees reported a wholesale shift from liquidity-taking to liquidity-making algorithms.

Conclusion

Let me be clear what this paper is not arguing. It is not claiming that humans and algorithms are identical beings: plainly they are not. Even in the brief narratives presented above, their different roles are clear. It is human beings, not algorithms, that

are angered by perceived queue jumping. It is humans, not algorithms, that are prosecuted for spoofing, and the traditional legal test – weakened by the Dodd-Frank Act, but still prominent in legal proceedings – is human intent: did Mr Sarao, for example, *intend* his orders to move prices?

Nevertheless, the previous sections of this paper have, I hope, shown that the limited forms of action available to trading algorithms (to submit orders, to cancel them, and sometimes to modify them) can nonetheless give rise to rich forms of strategic interaction. Algorithms use whatever means are made available to them to get to the front of the electronic queue; they dissimulate (sometimes legitimately, sometimes not); they seek to defend their processes of inference against the effects of dissimulation; some enjoy privileged powers denied to others. There is an increasingly strongly policed, but still vaguely defined, boundary between legitimate strategic action and illegal spoofing. As the boundary hardens, so the nature of strategic algorithmic action shifts.³ It is indeed perfectly possible that in the kinds of markets discussed here, algorithms now act *more* strategically than humans can.⁴ The very fact that human passions are raised by algorithmic queuing and spoofing, and that the latter can lead to jail, is indirectly testimony to the richness of how algorithms interact: we see in that interaction echoes of how we humans interact. As Knorr Cetina commented in response to a workshop presentation of this paper, the notion of ‘the interaction order of algorithms’ has a certain phenomenological adequacy.

³ The interviewee who told me about the decline in ‘classic’ forms of spoofing following the first criminal indictment also said that they were being replaced by orders that were still going to be cancelled, but acted ‘epistemologically’ (by revealing how ‘real’ other orders in the order book were) rather than by immediately profitable (but detectable and legally problematic) trades.

⁴ It is, for example, harder for a human spoofer (who can only act slowly) to hide his or her traces, for example, by using multiple small orders with random sizes, rather than a single all-to-obvious big order.

The brief discussion in the section immediately before this conclusion also demonstrates, I would argue, the relevance of one of the main reasons Goffman gave for 'isolating the interaction order': that it 'provides a means and a reason to examine diverse societies comparatively, and our own historically' (Goffman, 1983: 2). Look comparatively across asset classes (contrasting, for instance, foreign exchange and share trading), or historically examine how trading has changed, and you find in algorithmic interaction not just emergent phenomena, generated reflexively and locally, but also the traces of wider processes: the efforts to outlaw spoofing and thus keep order books 'pure'; the continuing market power of big banks in foreign exchange; even perhaps the decades-long neoliberal project to give competition – that unnatural, 'fragile' thing – a 'real, historical existence' (Foucault, 2008: 131-32).

Modesty, though, is also required, for by now the reader will surely have noticed a methodological irony. This paper has not employed the preferred methodology of interactionist sociology, participant observation. Remarkably, given that HFT firms protect their intellectual property fiercely (even gaining interview access is in many cases impossible), Robert Seyfert of the University of Konstanz and, especially, Ann-Christina Lange of the Copenhagen Business School have gained a degree of observational access to HFT firms (see Borch, Hansen and Lange, 2015). Observing an HFT firm, however, is not the same as observing algorithms. Algorithms were interacting in Cermak when I visited that datacentre, but were of course invisible to me. To be dependent, in consequence, on the testimony (or even to observe the actions) of the human beings who write and use trading algorithms is to rely upon indirect evidence that can mislead. As one of my HFT interviewees warned me: 'someone could be in all honesty saying they're [their algorithms are] doing [something] when in fact they're doing something else, they're just not measuring it right'.

The interaction of algorithms does leave its traces in changes in order books and in prices. However, in the order-book and price data available to academic researchers, trading-account identifiers are almost always removed, making it difficult or impossible to identify sequences of actions by the same algorithm or even the same trading firm. Researchers employed by regulatory bodies do have access to account identifiers, but they have found the task of unravelling patterns of algorithmic interaction (even in short time periods) computationally, and perhaps conceptually, close to intractable. Almost a decade on, there is still debate on the causes of the 'flash crash', a twenty-minute spasm in the US futures and stock markets on 6 May 2010. A working party from five regulatory bodies spent months seeking to disentangle a broadly similar event in the US Treasury bond market between 9:33 a.m. and 9:45 a.m. on 15 October 2014, but confessed themselves unable fully to identify '[t]he dynamics that drove ... trading' in those twelve minutes (Joint Staff Report, 2015: 33). Furthermore, any Goffmanian wants to see analyses of routine, not just unusual, interaction, but researchers employed by market regulators understandably often need to focus on the unusual.

We are, in short, still far from having a robust understanding of how trading algorithms interact. However, the virtue of the concept of 'interaction order' is that it focuses our attention on the right issue, which is indeed interaction. Any individual trading algorithm can perfectly reasonably be seen as the 'delegate' of a human being or beings (although my interviewee's warning of their possibly defective understanding of its operations must be borne in mind). But the ensemble of interacting algorithms is not our individual or collective delegate, and while the program text of a trading algorithm may usually remain unchanged by interaction, how it materially acts *is* shaped by interaction. Even individual algorithms thus need to be understood relationally, in the spirit of Goffman's unfortunately worded but succinct summary of his relational sociology: 'Not, then, men and their moments. Rather moments and their men' (Goffman, 1967: 3).

Appendix 1: Empirical Nuances

There are two other main ways (beyond the entry and cancellation of orders) in which trading algorithms interact. The first is in electronic trading venues that, unlike those discussed in the text, do not have central order books. For example, some venues (especially in bond trading and foreign exchange) have instead a fixed distinction between participants, generally algorithmic, that are allowed to post bids and offers – either in response to requests to do so, or in the form of constantly ‘streamed’ prices – and other participants, generally human beings, that cannot post prices but can only accept prices posted by others. Second, even although the different algorithms being run by a trading firm do not (as far as I can tell) usually collaborate, they can have effects on each other. Firms normally have aggregate risk limits that mean, e.g., that if one algorithm has built up a large position in a particular stock, others are prevented from adding to it. Also common is software to prevent self-trading (an algorithm selling financial instruments to another of the firm’s algorithms), which incurs unnecessary expenses and may attract unwelcome regulatory attention as potentially setting a ‘false price’. This software has the effect, e.g., that if one algorithm is offering shares at a given price, then (dependent on the software’s settings) either all of the firm’s other algorithms are prevented from bidding to buy shares at that price or the original offer is cancelled (in effect by an algorithm other than the one that submitted it).

Local interactions of this kind among algorithms are of theoretical interest, for example if one wishes to apply metaphors such as ‘swarming’. Also interesting in this respect is that algorithms can sometimes learn ‘locally’ about order-book changes via a ‘confirm’ – an electronic message reporting execution of one of a firm’s orders – before the corresponding message appears in the overall datafeed.

The discussion of orders in the text is also not exhaustive. As well as the orders described (which market participants would call 'limit orders', i.e. orders to buy at or below a specified price, or orders to sell at or above it), it is, for instance, also often possible for an algorithm or human trader to submit a 'market order'. This is an order simply to buy or to sell at the best available price, and it can therefore under almost all circumstances be executed immediately. The order book thus contains only limit orders, not market orders, which is why the fuller name for it is 'central limit-order book'.

While most matching engines operate a time-priority system of the kind described in the text, a minority employ 'pro-rata' matching, in which new executable orders are matched against existing orders in proportion to the size of the latter. Certain 'designated market makers' (for example in options) may also be guaranteed a specific proportion of any incoming executable order.

Appendix 2: Intermarket Sweep Orders and Special Order Types

Regulation NMS in effect decrees that before an exchange's matching engine adds a displayable order to its order book it must check the best (i.e. highest priced) bid and best (i.e. lowest priced) offer available at all other US exchanges. A matching engine cannot, for example, add to its order book an offer to sell at the price of the national best bid, but must electronically route that order for execution to the exchange whose book contains that bid. In consequence, incoming orders for shares are often delayed while the matching engine performs this check and waits for it to be permissible to add them to the order book.

Regulation NMS thus directly shapes queuing, and has given rise to a variety of ways in which algorithms can improve their queue positions. The most important and most prevalent is an exception in Regulation NMS (Securities and Exchange Commission, 2005: 37523) that provides for an 'Intermarket Sweep Order'. This is an

order bearing a computerised flag indicating that other orders have been sent to other exchanges that will execute against, and thus remove from their order books, any orders that block the addition of the flagged order to the order book. However, only registered broker-dealers and customers authorised by them are allowed to use the Intermarket Sweep Order flag.

Exchanges have themselves also designed new types of specialised orders, which often hinge on the fact that Regulation NMS governs the entry of displayable orders, not hidden orders. Matching engines always allocate hidden orders positions in the order-book queue behind displayable orders at the same price, but if those displayable orders cannot be added to the order book because of the constraints of Regulation NMS, an initially hidden order can still get to the head of the queue. Best known of these new orders is one made available by the exchange Direct Edge called 'Hide Not Slide' (Anon, 2009).

Acknowledgement

I am very grateful for research funding from the European Research Council (grant 291733) and UK Economic and Social Research Council (ES/R003173/1), as well as for helpful comments from *TCS*'s three referees.

References

- Anon. (2009) 'Direct Edge Launches Hide Not Slide Order, May 28', URL (consulted 31 December 2012): <http://www.thetradenews.com/print.aspx?id=2597>
- Arnoldi, J. (2015) 'Computer Algorithms, Market Manipulation and the Institutionalization of High Frequency Trading,' *Theory, Culture and Society*, early online.
- Beer, D. (2009) 'Power through the Algorithm? Participatory Web Cultures and the *Technological Unconscious*', *New Media & Society* 11(6): 985-1002.
- Bodek, H. (2013) *The Problem of HFT: Collected Writings on High Frequency Trading & Stock Market Structure Reform*. N.p.: Decimus Capital Markets.

- Borch, C., K. B. Hansen, and A.-C. Lange (2015) 'Markets, Bodies, and Rhythms: A Rhythmanalysis of Financial Markets from Open-Outcry Trading to High-Frequency Trading,' *Environment and Planning D: Society and Space*, early online.
- Bucher, T. (2012) 'Want to be on Top? Algorithmic Power and the Threat of Invisibility on Facebook', *New Media & Society* 14(7): 1164-80.
- Burns, T. (1992) *Erving Goffman*. London: Routledge.
- Cheney-Lippold, J. (2011) 'A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control', *Theory, Culture and Society* 28(6): 164-81.
- Deleuze, G. (1992) 'Postscript on the Societies of Control', *October* 59(Winter): 3-7.
- Foucault, M. (2008) *The Birth of Biopolitics*. London: Palgrave Macmillan.
- Gane, N., V. Couze, and M. Hand (2007) 'Ubiquitous Surveillance: Interview with Katherine Hayles', *Theory, Culture and Society* 24(7-8): 349-58.
- Goffman, E. (1959) *The Presentation of Self in Everyday Life*. New York: Doubleday.
- . (1963) *Behavior in Public Places: Notes on the Social Organization of Gatherings*. New York: Free Press.
- . (1964) 'The Neglected Situation', *American Anthropologist* 66(6): 133-36.
- . (1967) *Interaction Ritual: Essays on Face-to-Face Behavior*. New York: Pantheon.
- . (1968) *Stigma: Notes on the Management of Spoiled Identity*. Harmondsworth, Middlesex: Penguin.
- . (1983) 'The Interaction Order', *American Sociological Review* 48(1): 1-17.
- Hayles, N. K. (1999) *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press.
- . (2006) 'Unfinished Work: From Cyborg to the Cognisphere', *Theory, Culture and Society* 23(7-8): 159-66.
- . (2012) *How We Think: Digital Media and Contemporary Technogenesis*. Chicago: University of Chicago Press.
- Joint Staff Report. (2015) 'The U.S. Treasury Market on October 15, 2014', U.S. Department of the Treasury, Board of Governors of the Federal Reserve System, Federal Reserve Bank of New York, U.S. Securities and Exchange Commission and U.S. Commodity Futures Trading Commission, 13 July, URL (consulted 20 August 2015): http://www.treasury.gov/press-center/press-releases/Documents/Joint_Staff_Report_Treasury_10-15-2015.pdf
- Kittler, F. (2006) 'Lightning and Series – Event and Thunder', *Theory, Culture and Society* 23(7-8): 63-74.
- Knorr Cetina, K. (2009) 'The Synthetic Situation: Interactionism for a Global World', *Symbolic Interaction* 32(1): 61-87.
- . (2013) 'Presentation to Panel, Theorizing Numbers', in *American Sociological Association Annual Meeting*. New York.
- Knorr Cetina, K. and U. Bruegger. (2002a) 'Global Microstructures: The Virtual Societies of Financial Markets', *American Journal of Sociology* 107: 905-51.
- . (2002b) 'Traders' Engagement with Markets: A Postsocial Relationship', *Theory, Culture and Society* 19(5-6): 161-85.
- Knorr Cetina, K. and A. Preda. (2007) 'The Temporalization of Financial Markets: From Network to Flow', *Theory, Culture and Society* 24(7-8): 116-38.
- Lash, S. (2002) *Critique of Information*. London: Sage.
- . (2007) 'Power after Hegemony: Cultural Studies in Mutation', *Theory, Culture and Society* 24(3): 55-78.
- Latour, B. (2005) *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.
- Lewis, M. (2014) *Flash Boys: Cracking the Money Code*. London: Penguin.
- MacKenzie, D. (2014a) 'Be Grateful for Drizzle', *London Review of Books* 36(17): 27-30.

- . (2014b) 'At Cermak', *London Review of Books* 36(23): 25.
- Pardo-Guerra, J.P. (Forthcoming) *Orders of Finance*. Cambridge, MA: MIT Press.
- Parrika, J. (2015) *A Geology of Media*. Minneapolis, MN: University of Minnesota Press.
- Preda, A. (2009) 'Brief Encounters: Calculation and the Interaction Order of Anonymous Electronic Markets', *Accounting, Organizations and Society* 34: 675-93.
- . (2013) 'Tags, Transaction Types and Communication in Online Anonymous Markets', *Socio-Economic Review* 11: 31-56.
- Savat, D. and M. Poster (Ed.) (2005) *Deleuze and New Technology*. Edinburgh: Edinburgh University Press.
- SEC (Securities and Exchange Commission). (2005) '17 CFR Parts 200, 201, et al: Regulation NMS; Final Rule', *Federal Register* 70 (124): 37496-644.
- US Department of Justice. (2015) 'United States of America v. Navinder Singh Sarao: Criminal Complaint', United States District Court, Northern District of Illinois, Eastern Division, 15 CR 75, URL (consulted 19 August 2015): <http://www.justice.gov>
- Vehlken, S. (2013) 'Zootechnologies: Swarming as a Cultural Technique', *Theory, Culture and Society* 30(6): 110-31.
- Zaloom, C. (2006) *Out of the Pits: Trading and Technology from Chicago to London*. Chicago: Chicago University Press.

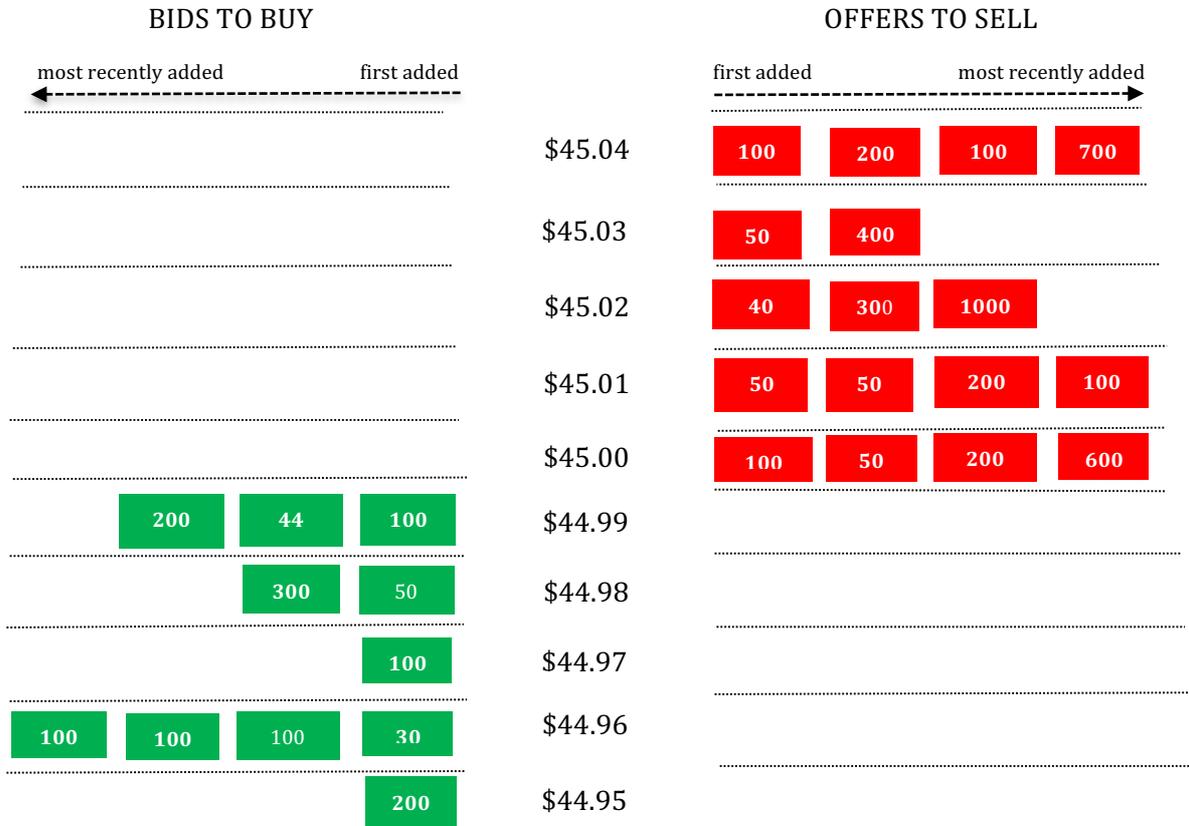


Figure 1: An example of an order book