Machine learning and social theory: Collective machine behaviour in algorithmic trading

Christian Borch

Copenhagen Business School, Frederiksberg, Denmark

Abstract

This article examines what the rise in machine learning (ML) systems might mean for social theory. Focusing on financial markets, in which algorithmic securities trading founded on ML-based decision-making is gaining traction, I discuss the extent to which established sociological notions remain relevant or demand a reconsideration when applied to an ML context. I argue that ML systems have some capacity for agency and for engaging in forms of collective machine behaviour, in which ML systems interact with other machines. However, ML-based collective machine behaviour is irreducible to human decision-making and thereby challenges established sociological notions of financial markets (including that of embeddedness). I argue that such behaviour can nonetheless be analysed through an adaptation of sociological theories of interaction and collective behaviour.

Keywords

Algorithmic trading, collective behaviour, embeddedness, interaction, machine learning

In a multi-authored Nature manifesto on ‘machine behaviour’, Rahwan et al. (2019) propose a research agenda that investigates machines powered by artificial intelligence (AI), including sophisticated machine learning (ML) technologies. As the authors note, these types of technologies are profoundly reshaping society and, therefore, demand...
careful scrutiny across several disciplines. Rahwan et al. (2019) outline three particular scales of inquiry that call for interdisciplinary research: individual machine behaviour (the study of individual algorithms), human–machine behaviour (humans interacting with machines), and collective machine behaviour (machines interacting with other machines). Although individual machine behaviour is still a topic primarily studied by computer scientists, sociologists and others have made significant contributions especially to the investigation of human–machine interactions in the ML domain. This includes studies of the kinds of class, gender, race and other biases that ML systems may produce and reproduce (e.g. Brayne & Christin, 2020; Buolamwini & Gebru, 2018; Eubanks, 2018; Noble, 2018), just as scholars have discussed the broader ethical and political implications of ML systems (e.g. Amoore, 2020; Aradau & Blanke, 2017; Coeckelbergh, 2020). These are crucial topics to examine and deserve further scrutiny as ML systems are rolled out in ever-more domains. What has received less attention is what ML might mean for sociological theory, including whether established sociological concepts remain relevant in an ML era or whether a substantial rethinking of conventional notions is needed in light of how ML works and affects society (Edelmann et al., 2020).

Against this background, this article aims to discuss some of the implications that ML systems, and the ensuing collective machine behaviours they engage in, have for social theory. I suggest that the task of reflecting upon the extent to which existing sociological notions – developed primarily for understanding a society inhabited by human beings – continue to have analytical purchase when it comes to understanding ML entails a critical as well as a (re)constructive side. Consequently, I critically examine several existing sociological notions considering ML systems, demonstrating their limitations when it comes to understanding how these kinds of systems work. However, I also seek to develop, in a more reconstructive fashion, ways in which sociological notions can be revised and deployed to provide insight into ML systems. Specifically, I focus on the notions of interaction and embeddedness (both of which are key to existing debates about financial markets) and collective behaviour (as per the interest in collective machine behaviour).

I make two overall arguments. First, sociology should take the capability of ML systems to make autonomous decisions seriously. Although this does not encompass consciousness, intentionality or meaning in any recognizable human form, it does suggest that ML systems has some capacity for agency. This sets ML systems apart from non-ML-based, human-defined algorithmic systems – whereas non-ML-based algorithmic systems can be seen as a direct extension of human decision-making, ML systems cannot. Second, ML changes the field of inter-machinic connections, or what Rahwan et al. (2019) refer to as collective machine behaviour. While sociologists have started pondering the theoretical implications of human–machine interactions involving ML systems (e.g. Esposito, 2017; Fourcade & Johns, 2020; see further Brighenti & Pavoni, 2021; Wagner-Pacifici et al., 2015), collective machine behaviour that involves ML systems remains largely unexplored from the perspective of social theory. As Edelmann et al. remark, ‘Sociologists have […] left mostly untheorized the ways that interaction between machines themselves can shape patterns of human behavior and social organization as well’ (2020, p. 74). Therefore, this article pays particular attention to this level
of collective machine behaviour which, given the ML systems’ capacity for agency, is irreducible to human decision-making.

I contend that automated securities trading is a relevant field to study inter-machinic connections and collective machine behaviour, although I believe that the theoretical reflections presented in this article have wider applications and capture the role of ML systems in non-financial domains as well. Briefly, the evolution of securities trading looks like this: In the past, decisions to buy or sell securities were made by human traders flocking on exchange trading floors. Alongside the increasing computerization of society, and at a particular speed since the early 2000s, human traders have been progressively replaced by fully automated computer algorithms; thus, in many markets today, most buy and sell orders are sent by the so-called high-frequency trading (HFT) algorithms (MacKenzie, 2021). These are fully automated, high-speed algorithms that place orders without human intervention, and which typically respond to market developments within microseconds, that is, millionths of a second (Lange, 2016; Lange et al., 2016; MacKenzie, 2018a, 2018b, 2021; Pardo-Guerra, 2019; Seyfert, 2016; see also Svetlova, 2021). Such algorithms are conceived by humans and, thus, act out human intentionality. However, during the past 5 years, ML systems have been gaining traction in financial markets, including for automated trading purposes, and this has given rise to trading strategies that are entirely machine generated (Guida, 2019; Hansen, 2020, 2021; Hansen & Borch, 2021; López de Prado, 2018).

I begin the analysis by briefly explaining what ML is and how it is deployed in securities trading. I particularly emphasize the autonomous decision-making ability of ML systems vis-à-vis humans and contrast ML systems with their non-ML forerunners, that is, human-defined algorithmic systems whose rules and decisions are not automatically generated but act out human intentionality. I then discuss ML’s social theory implications by focusing on collective machine behaviour in which ML systems interact with other machines. This discussion is divided into three sections that revolve around interaction, embeddedness and what I call distinctly collective dynamics, respectively. The conclusion summarizes.

What is ML and how is it used in securities trading?

ML is a subset of AI. The latter may be defined as ‘intelligence displayed or simulated by code (algorithms) or machines’ (Coeckelbergh, 2020, p. 44), where intelligence needs not be modelled exclusively around human intelligence. Broadly speaking, ML is concerned with learning from data to extract patterns and make predictions. For example, ML is deployed to detect malware, just as companies such as Netflix use it to recommend shows and films to customers. Less quotidian examples include the AlphaGo program that, based on a complex neural network architecture, achieved superhuman performance by learning from playing against itself; this learning was not accomplished through human guidance or founded on human knowledge (Silver et al., 2017). The AlphaGo example captures the basic idea of ML, where an algorithmic system can be trained on a large set of data (input), teaching it to produce a certain output, typically in the form of an action-oriented prediction. The learning phase can be designed differently, and there is a
host of ML techniques available depending on the particular learning task and the desired output type.

Some trading firms, which specialize exclusively in ML-based systems, use deep neural networks. In this ML architecture, initially loosely built upon the idea of the human brain, learning is achieved by training layers of information processing units, neurons, to extract increasingly abstract features from data (for an introduction, see Kelleher, 2019). Other firms would use other ML techniques or various combinations of ML approaches. For the argument of this article, the specific kind of ML architecture is less important. What is crucial, instead, is that for all these firms, their ML system is given an objective function and a lot of data to automatically learn how to extract patterns from these data. On this basis, the ML system automatically devises a trading strategy which is then automatically implemented in the markets. The data these systems are being fed usually consist of ‘electronic order-book’ data, that is, securities exchanges’ electronic registration of all pending orders to buy or sell securities (Pardo-Guerra, 2019).

Non-ML-based, human-defined trading systems also rely on order-book data. However, the central difference to ML systems is that for the non-ML systems, the algorithmic strategy would be designed by one or more human traders (or ‘quants’), translated into code, and then automatically executed by the algorithmic system. In other words, each part of the trading strategy – its rules – would be conceived by humans and coded by hand, and therefore the algorithmic implementation would, in principle, be a purely machinic enactment of the human-defined strategy. A rule is an instruction given to an algorithm outlining how it should behave in a particular way under given circumstances. With a hypothetical example, a rule may involve placing a buy order for a certain number of Apple shares if the price of these shares falls by 0.2 per cent. To determine whether to place this order, the trading algorithm monitors price movements as recorded in the electronic order book (the rule is applied to an ongoing flow of incoming data). In contrast, the idea of ML-based trading is that the ML system comes up with the rules itself depending on its objective function and the data it is fed. So, here, rules or strategies are not conceived by humans but rather by the ML system itself.

I argue that this difference is highly consequential and has important social theory implications. Although humans design a particular ML architecture – and this can be an immensely labour-intensive process – humans are, in effect, removed from the equation when it comes to identifying and implementing trading rules. This does not exclude several human influences, but it is important to understand their roles properly. Humans define the ML system’s objective function and decide what data the ML systems are fed. Humans also train the ML system as well as they design the overall algorithmic trading infrastructure. That said, it is important to appreciate that for firms specializing in ML-based trading, humans do not directly affect or intervene in the trading policies generated by the ML systems. Rather, in deep neural network models, for example, humans might introduce certain incentives in the form of penalties and rewards, but the neural network would be tasked with producing its own rules.

I suggest that Luhmann’s (1995) sociological systems theory offers a helpful way of conceiving this relationship between human input and automated rule generation and decision-making. Mobilizing Luhmann for understanding algorithmic systems is not
new. For example, Esposito (2017) argued for analysing AI systems on the basis of Luhmann’s systems theory, emphasizing their ability to engage in communication and seeing this as more important than debating whether these systems possess intelligence or not. The key analytical benefit of Luhmann’s approach is, according to Esposito, that it detaches communication from psychological consciousness, treating each as an independent domain.

Underpinning this notion is Luhmann’s insistence that both social and psychic systems are autopoietic, operationally closed systems; that is, they produce and reproduce their own constituting elements and maintain a strict operational boundary to their environment. However, even though operations are system-specific and cannot cross system boundaries, operational closure does not imply empirical isolation. A system may depend on particular material conditions in its environment (e.g. nourishment or data), just as operational closure enables systems to be cognitively open towards their surroundings (e.g. by registering what takes place in the environment or in the electronic order book). That said, given that systems are operationally closed, System A cannot interfere directly with the operations of System B. At most, System A can produce an external ‘irritation’ of System B, but the effect, if any, that is generated in System B is entirely up for the latter to decide (Luhmann, 2012, p. 56).

This operational closure captures the relationship between humans and their ML systems. Once ML trading systems are up and running, human influence is reduced to an irritation between operationally closed, autopoietic systems; humans may make their complexity (their understanding and knowledge of markets) available to the ML systems, but, eventually, the latter decides which irritations are relevant to them and whether and to what extent these irritations should affect their trading rules and behaviours. This, again, sets them apart from non-ML-based, human-defined trading algorithms, which, since they perform a direct enactment of human-conceived rules, are characterized by no operational closure vis-à-vis their human inventors.

It is important to be clear about what this argument about ML systems’ independent decision-making capacity means. To repeat, it is not a claim about ML systems being entirely autonomous. As mentioned, ML systems operate against a backdrop of considerable human curation. Further, ML trading systems cannot suddenly jump beyond the types of data on which they are trained (just as an image-recognition ML system cannot immediately be used for speech recognition). For example, an ML system trained on data from one type of market regime (e.g. pre-financial crisis data) may have difficulties making good predictions when deployed in a drastically different market setting (e.g. post-2008). In other words, there are several limitations to the kind of agency an ML trading system may have. However, what is distinctly new to these systems compared to their human-defined forerunners is that the actual trading rules/strategies are entirely machine generated. This does not rule out that ML trading systems pick up patterns in data that reflect human decision-making. For example, it is conceivable that an ML trading system might identify patterns in order-book data that reflect the ways in which human-defined algorithms are trading – and then comes up with a trading strategy that considers such human-defined decision-making. Still, the extraction of such patterns and how to develop a trading strategy out of them is the task for the ML system itself.
I suggest that the operational autonomy of ML trading systems – their ability, once designed, to extract patterns independently and devise autonomous trading strategies – is a relevant starting point for reflecting upon what social theory implications these kinds of systems may have and how to conceive of such systems theoretically. There may be several relevant theoretical routes to follow here. For example, Latour’s (2005) work and broader actor–network theory arguments about non-human actants may seem an obvious point of departure. However, given that Latour’s interest essentially revolves around the ways in humans and non-humans interact, and since, as mentioned earlier, in this article, I am primarily interested in inter-machinic connections, I propose to centre the discussion on the notion of collective behaviour (about which Latour has less to say). This is particularly relevant in the context of automated trading, as ML-based trading systems never act in a vacuum. Per definition, trading involves more parties, and present-day financial markets are made up by numerous automated trading systems interacting with one another (MacKenzie, 2019, 2021). Since the ML-based trading policies are machine-generated, the ensuing collective behaviour is irreducible to human decision-making. This raises important questions about how to sociologically theorize such collective machine behaviour.

**Collective machine behaviour**

According to the definition proposed by Rahwan et al. (2019), ‘the study of collective machine behaviour focuses on the interactive and systemwide behaviours of collections of machine agents’ (p. 482). Perhaps reflecting the fact that the authors take inspiration from work on animal behaviour rather than social theory, Rahwan et al. do not spell out what ‘interactive and systemwide’ behaviours might entail, although they do note that:

> human-made AI systems do not necessarily face the same constraints as do organisms, and collective assemblages of machines provide new capabilities, such as instant global communication, that can lead to entirely new collective behavioural patterns. Studies in collective machine behaviour examine the properties of assemblages of machines as well as the unexpected properties that can emerge from these complex systems of interactions. (2019, p. 482)

From a sociological perspective, the emphasis on collective machine behaviour is at once important and underspecified. It is important because it pinpoints a level of interaction that is both potentially highly consequential (think of a conflict between automated warfare systems) and theoretically challenging given that it unfolds between non-human agents. However, the reference to collective machine behaviour is also underspecified because Rahwan et al. do not differentiate between different forms of collective behaviour. In contrast, sociological theory is rich in perspectives that detail various types of collective behaviour. However, since these theories were developed to capture inter-human collective behaviours, it is not obvious that they would also be relevant for theorizing collective machine behaviour. I discuss this potential applicability below by attending to how three levels of collective machine behaviour could be analysed in the field of algorithmic trading.


**Interaction**

Interaction builds on an analysis by MacKenzie (2019) in which he mobilizes Goffman’s (1983) notion of the ‘interaction order’ to shed light on the interaction dynamics of fully automated trading algorithms. While appreciating that applying the sociological face-to-face approach par excellence to non-human systems might seem surprising, MacKenzie stresses that a similar analytical path has been trodden by scholars such as Knorr Cetina and Preda, both of whom productively extended Goffmanian insights to the study of (non-automated) electronic markets (Knorr Cetina, 2009; Knorr Cetina & Bruegger, 2002; Preda, 2009, 2017). Building on this research, MacKenzie argues that, in the context of algorithmic trading, a distinct Goffmanian interaction order may, for example, be identified when automated trading algorithms pursue particular forms of queuing practices (focusing on how they place orders in the electronic order book) as well as when they engage in dissimulation – such as when an algorithm seeks to manipulate markets by placing orders that give the impression of a larger market movement and then take advantage of how other market participants might be fooled by this strategy (see also MacKenzie, 2018a). For example, a manipulative algorithmic strategy seeking to sell an accumulated quantity of shares to the best possible price might consist in placing a large number of orders to buy that particular stock, hoping to ignite other algorithms to increase their sell price and then sell at the better price while quickly cancelling its orders to buy.

Through examples like this, MacKenzie asserts that ‘the limited forms of action available to trading algorithms [. . .] can nonetheless give rise to rich forms of strategic interaction’ (2019, p. 53). Therefore, although there are obvious differences between humans and algorithms, it is reasonable to see the latter as an extension of the former while simultaneously recognizing that algorithmic interactions assume a distinct order that is irreducible to what the human inventors had in mind. In MacKenzie’s words:

> Any individual trading algorithm can perfectly reasonably be seen as the ‘delegate’ of a human being or beings [. . .]. 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. (2019, p. 55, original emphasis)

I would modify the first of these claims. It is true that human-defined algorithms can be seen as human delegates, although, as MacKenzie acknowledges, the people who are developing and deploying human-defined algorithms might not always have full purview over the ways they act in markets. However, for ML algorithms, it is misleading to consider them human delegates since this fails to appreciate their independent rule-generating and decision-making capabilities.

That said, MacKenzie is correct in his second claim that the interactions between fully automated trading algorithms will often affect how individual algorithms act. This holds for human-defined and ML-based trading algorithms alike. For a human-defined algorithm, this is because it is programmed to observe and respond to order-book dynamics, meaning that when it responds to one market movement by sending, modifying or
cancelling orders, this creates a change in the electronic order book that may trigger other algorithms to react, generating new changes to which it responds and so on. The same applies to ML trading algorithms but, additionally, interactions will also lead to material changes in their operations because they learn from these interactions. In other words, the behaviour of the individual ML algorithms would change through learning, as they continuously adapt their rules and behaviours in light of their interactions with other algorithms. This suggests that algorithmic interaction involving ML-based algorithms constitutes a potentially more dynamic interaction order than the one MacKenzie identifies among human-defined algorithms and one that is, therefore, even closer to what Goffman analysed. Of course, this should not lead to a disregard of the substantial differences between humans and machines. That said, MacKenzie’s point that algorithmic interaction may be analysed according to Goffman’s work seems even more appropriate in the context of ML-based trading algorithms.

A central feature of any interaction order is that it is founded on repeated interactions. Although an interaction order may affect exchanges between participants who have never interacted before, the order itself can only arise based on past recurrent encounters. In addition, the interaction order may be influenced by social relationships and social structures that exist outside it (Goffman, 1983) – including, in the field of algorithmic trading, regulation and the ways in which securities exchanges organize trading (MacKenzie, 2019, pp. 51–3; Pardo-Guerra, 2019). Acknowledging the role of repeated interactions and social relationships allows for an additional set of social theory considerations that go beyond a Goffman-inflected analysis of the interaction order and help to further specify what collective machine behaviour might involve from a sociological point of view. Specifically, in their examinations of human traders on traditional trading floors, Abolafia (1996) and Baker (1984) demonstrated that the repeated interactions of these traders generated a normative order in which their economic action was rooted. This meant that if a market participant violated the norms, peers might use various tactics to trade with others instead. The observations by Abolafia and Baker offer empirical support, in the context of securities trading, of Granovetter’s (1985) notion of embeddedness. The question this kind of research prompts is whether embeddedness remains analytically relevant in an era of ML-based algorithmic trading. If it is, this would suggest that collective machine behaviour may not just be analysed in terms of algorithmic interactions (as per MacKenzie’s analysis) but also – and more specifically – according to how repeated interactions may create forms of embeddedness that co-shape the interaction patterns.

**Embeddedness**

Granovetter’s general argument is that, when considering human economic activity, it is crucial to acknowledge that, ‘actors do not behave or decide as atoms outside a social context, nor do they adhere slavishly to a script written for them [. . .]. Their attempts at purposive action are instead embedded in concrete, ongoing systems of social relations’ (1985, p. 487). In other words, human economic actors neither act in a social vacuum, nor do they mechanically enact particular norms or roles. Rather, their behaviour is rooted in and shaped by the ongoing social relations they maintain with others. This

\[\text{Equation}\]

\[\text{Equation}\]
is, to repeat, precisely what Abolafia and Baker found. In the past, the human traders populating the trading floors did not just pursue narrow economic interests. Their ongoing social relationships – the fact that they spent years trading with each other on the floors – also produced informal norms about proper trading conduct which regulated their economic interactions.

Does this notion of the social embeddedness of economic action remain relevant for an era of securities trading in which interaction plays out among fully automated algorithms rather than human beings? However unlikely this might seem, Granovetterian embeddedness continues to have analytical purchase in two dimensions. The first concerns those *inter-human* interactions that remain important within algorithmic trading. Firms specializing in both human-designed and ML-based algorithmic trading consider it crucial to entertain ongoing social relationships especially with exchange staff and third-party data providers (see also MacKenzie, 2021). This includes keeping track of new exchange updates such as new order types. While such inter-human relationships do not affect trading directly, they are considered important for overall trading operations and therefore constitute a central wrapper around automated markets. Similarly, human staff members interact within individual trading firms, meaning that intra-organizational work may be socially embedded, although it is not uncommon to maintain a high degree of secrecy even within firms in the industry (Lange, 2016; see also Souleles, 2019).

The next dimension concerns *human–machine* interactions. It is common for the people who design and develop algorithmic trading systems to become emotionally attached to them (Borch & Lange, 2017), although this kind of attachment is one-sided from human to machine rather than reciprocated. In Granovetter’s terms, these ways of interacting with machines might appear as a way of mimicking ongoing inter-human relations and forming a basis on which a kind of surrogate embeddedness might manifest, although, again, it is one-sided rather than reciprocated.

Still, arguably the most theoretically thought-provoking dimension concerns that of *inter-algorithmic* relations, which refer to the possibility of fully automated algorithms being socially embedded with one another. What might point towards the possibility of a social embeddedness of algorithms is their repeated interactions since repeated interactions form the condition of possibility of Granovetterian embeddedness. Indeed, repeated interactions might well take place in algorithmic trading. It is possible, for example, that algorithm \(X\) repeatedly interacts with algorithm \(Y\) (rather than with, say, algorithms \(A\), \(B\) or \(C\)) in the electronic order book. That said, what speaks strongly against the notion of ML algorithms being embedded in social relations with each other is a basic feature of automated trading that sets it apart from earlier modes of trading. Contrary to present-day markets, previous physical exchange trading floors were non-anonymous. Traders knew each other’s identity and often spent time together after work; this was part of what ensured their social embeddedness. By contrast, automated markets are anonymous as a rule, meaning that no market participant who enters an order in today’s electronic markets knows for sure who their counterparties are. Algorithms might detect that other particular algorithms follow specific patterns but which firms are behind these algorithms is generally not a piece of information they have access to (MacKenzie, 2018a). It follows that, although particular algorithms might be interacting with one another
repeatedly, their lack of knowledge about counterparty identity renders Granovetterian social embeddedness impossible among them.

Granovetter’s notion of social embeddedness is based on what might be called a ‘deep’ or ‘warm’ network conception of sociality. The more economic actors know one another, and the deeper their concrete personal relationships run, the more their ‘continuing economic relations [. . .] become overlaid with social content that carries strong expectations of trust and abstention from opportunism’ (Granovetter, 1985, p. 490). While the absence of such relations in the inter-algorithmic trading space means that, unlike their human-dominated forerunners, today’s financial markets are not socially embedded, this need not preclude the existence of sociality – or ‘social content’, as Granovetter calls it – among automated trading algorithms. Such sociality simply materializes in colder and more shallow forms. For example, elsewhere it has been suggested that Simmel’s (1997, p. 183) reflections on the particular urban sociality arising out of ‘brief metropolitan contacts’ – fleeting meetings between urban inhabitants on the streets or in the subway – may capture how automated trading algorithms interrelate in the electronic order book (Borch, 2020). I add to this the conjecture that the turn to ML-based algorithmic trading, in which algorithms have the capacity to learn from their interactions in the market, might pave the way for some middle-ground sociality between fleeting Simmelian encounters and deep Granovetterian social content. This takes us back to Goffman’s interaction order, which could be seen as constituting such a middle ground. That said, as Simmel (1992) observed, interactions might occasionally erupt into uncontrollable crowd and collective behaviour – and this goes, as I argue, beyond what Goffman’s notion captures.

**Distinctly collective dynamics**

When discussing collective machine behaviour, Rahwan et al. (2019) refer to ‘flash crash’ events – sudden extreme market movements that arise out of the interactions among automated trading algorithms – as a core illustration of what such behaviour might entail and how it may manifest. ‘Flash crashes are examples of clearly unintended consequences of (interacting) algorithms; leading to the question of whether algorithms could interact to create a larger market crisis’ (p. 483). Debates about algorithmic flash crashes often focus on the Flash Crash of 6 May 2010, a dramatic unanticipated market shock that generated losses in the range of USD 1 trillion in less than 30 min before trading was suspended (markets quickly recovered when trading resumed). Although the causes of the Flash Crash remain unclear (Borch, 2016; MacKenzie, 2021), it is generally acknowledged that the event was exacerbated because of the interactions between fully automated algorithms and because many market participants withdrew from markets as prices fell sharply, further aggravating the crash.

Other flash crashes have been identified and discussed since 2010. In fact, scholars have argued that smaller flash crashes occur frequently, with US data suggesting an average of about 14 such events per day (Golub et al., 2012; Johnson et al., 2013). According to some of this research, flash crash events are examples of what might happen once algorithms engage in ‘crowding’ and ‘herding’, quickly imitating one another (Johnson et al., 2013; MacKenzie, 2021). For example, Sornette and von der
Becke argue that, in automated markets, there is a risk that several algorithms ‘crowd to the same signal [i.e., tradable piece of information, typically from the order book]’, which can ‘lead to transient large instabilities’ where algorithms ‘form large destabilizing crowds’, and where flash crash events are an illustration of ‘panic herding with procyclical mutual excitations between HFT and the rest of the investor population’ (2011, pp. 7, 11, 12).

It is important to note that flash crashes need not emerge exclusively because of collective machine behaviours. For example, it has been argued that the distressing effects of the 2010 Flash Crash were exacerbated because some key market data were delayed during the event, which prompted ‘automated data-integrity checks’ of trading algorithms to make them shut down (MacKenzie, 2021: 229; see also Aldrich et al., 2017; Min & Borch, 2021). In other words, the interaction effects of algorithms may be conditioned and further aggravated by other factors. Moreover, although flash crashes are often highlighted as central illustrations of collective machine behaviour in markets, such behaviour may also materialize in non-flash crash situations. In addition to the examples given earlier (including that of market manipulation), collective machine behaviour may arise when several brokers acting on behalf of investor clients execute their orders at the same time. They would typically do so by breaking up large orders ‘into smaller slices proportional in size to the fluctuating overall volume of trading in the stock in question’ (MacKenzie, 2021, p. 231). As MacKenzie notes, this strategy works fine if there is only one algorithm doing this at the same. However, when several independent execution algorithms are doing this simultaneously, they are likely to adversely affect one another and generate unanticipated collective interaction dynamics.

The central point here is that collective machine behaviours involving automated trading systems may manifest in several ways, including when they seek to track competing algorithms, infer their strategies and trade against them; when they respond to the same information in the order book, potentially leading to widespread herding; and when complex feedback loops arise because the orders sent by one (set of) algorithms are used as input by others when placing their orders.

The possibility of crowd and herd dynamics emerging out of the interactions among fully automated trading algorithms suggests a need for rethinking the sociological vocabulary with which to understand collective machine behaviour. Specifically, this level of collective behaviour is difficult to appreciate based on either Goffman’s notion of the interaction order or Granovetterian embeddedness. MacKenzie (2019) acknowledges that Goffman’s work may be of little analytical use for understanding such dynamics, as Goffman was mainly interested in routine activities rather than outbursts of collective behaviour. Indeed, Goffman stresses that his analytical interest in ‘the study of ordinary human traffic’ marks the flipside to examinations of collective behaviour in the form of ‘riots, crowds, [and] panics’ (1963, p. 4). Importantly, Goffman’s concession about the analytical division of labour, between collective behaviour theory and his own approach, suggests that the latter is ill-suited when it comes to understanding collective machine behaviour among fully automated trading algorithms (be they human-defined or ML-based).

Does the broader sociological tradition of collective behaviour theory have more to offer? Providing a clear answer to this question is difficult given that this tradition is far
from homogeneous, stretching from classics such as Tarde (1989), Durkheim (1995), Park (1972) and Blumer (1939) to the collective behaviour theorization beginning in the late 1950s (Lang & Lang, 1961; Smelser, 1962; Turner & Killian, 1957) and extending into the social movement studies tradition into which the examination of collective behaviour was eventually (partially) absorbed (Tilly, 1978, 2004). At the risk of oversimplifying matters, I suggest that the main notion characterizing the post-1960s theorization – what might be termed ‘modern collective behaviour theory’ to distinguish it from its ‘classical’ forerunners – can be formulated as follows (Borch, 2006): Instead of seeing collective behaviour as driven by collective sentiments and irrational group unconsciousness, which might materialize in violent outbursts (ideas all associated with the classics), sociologists now portrayed collective behaviour as the aggregate action of rational individuals each pursuing just objectives (e.g. Berk, 1974; McPhail, 1991).

Evidently, modern collective behaviour theory was not conceived to explain algorithms. However, its emphasis on the rational behaviour of individual agents does have some salience when it comes to understanding automated markets. Human-defined algorithms are invented and designed to pursue individual strategies that make money. Although many firms would deliberately let some of their algorithms lose a bit of money (e.g. to tease out information about larger market movements), or accept that even key algorithmic strategies would occasionally be in the red, the entire rationale of this form of trading is to develop algorithms that are overall profitable. Similarly, the basic idea behind the algorithmic architectures of ML-based trading systems is to automate the process of creating rational, profit-optimizing strategies.

While the micro-foundation of modern collective behaviour theory – its emphasis on rational individual agents – aligns with algorithmic trading, I am less convinced about its usefulness when it comes to understanding the collective effects of algorithmic interactions. In human collective behaviour, people might either share the same objective or a common goal would assimilate their individual interests. Given that algorithmic trading is anonymous, algorithms are unaware of each other’s specific strategic interests. While some degree of imitation between especially human-designed algorithmic strategies is conceivable because there is some circulation of employees (and therefore ideas) between trading firms, nothing points towards the fact that individual algorithms – neither human-defined nor ML-based ones – submit to some common goal.

A more fundamental problem with modern collective behaviour theory subscribing to the notion of rational agents pursuing their individual interests is that it has little to say about distinctly collective dynamics, that is, collective events that are irreducible to individual action (Borch, 2020). Since collective behaviour is seen in this tradition as a mere numerical aggregation of individuals, without any extra-individual dynamism emerging out of their agglomeration, it is not significant for understanding flash crash events and similar unanticipated collective effects of algorithmic interactions in markets. Given this, I suggest that the classical tradition of collective behaviour theorization may have greater analytical relevance when it comes to understanding collective machine behaviour. The point here is not to say that this tradition should be resuscitated in its entirety. It has many problematic aspects – including sometimes operating with gendered and racial biases (this applies, in particular, to Le Bon, 1960) – which is why modern collective behaviour theory emerged as a critical response to it. However, I maintain that
it is possible to avoid these biases and revive elements from the classical tradition in a present-day analytical context.

This includes the notion that collective behaviour is irreducible to individual action. It is not possible, scholars within this tradition have argued, to infer from micro-level assumptions about agent behaviours to the collective dynamics in which the individuals take part. Rather, to put it in later cybernetic language, some emergent or self-organizing properties should be acknowledged in collective behaviour. This is what Tarde (1968, p. 323) tried to establish when describing collective behaviour as ‘a sudden organization, a spontaneous generation’ in which ‘incoherence becomes coherence’ and ‘noise’ transforms into ‘voice’. Durkheim (1995) was on a similar track when portraying collective effervescence as something that momentarily transforms the individual, pointing to a new form of sociality. Simmel (1950, p. 93), too, can be included here, seeing collective behaviour as an event that ‘overwhelms the individuals’ and carries them away against their intentions. I am not suggesting that these notions can be directly transplanted onto the field of automated trading. For example, central to the accounts of Durkheim, Simmel and Tarde is the role of emotions and passions, which finds no equivalent in non-human, fully automated trading algorithms. Still, what this type of collective behaviour theory suggests is that the interaction of agents can produce an unanticipated surplus effect – that something new and extra arises out of the interactions of individual agents.

This may be further spelled out in at least two ways. First, collective machine behaviour need not entail ongoing relationships among algorithms. Although certain conditions might be more likely than others to trigger particular types of collective machine behaviour (and better understanding such conditions is important) what is central about the latter is that they can erupt suddenly and unexpectedly, being activated by small individual actions or by some combination of minor events. This is intensified, second, by the fact that ML-based algorithms come up with their own trading policies and implement these automatically, meaning that any collective machine behaviours are thus less foreseeable from a human point of view. In fact, given that, in particular, neural network-based ML architectures are generally associated with opacity – even ML experts struggle to understand how they arrive at their predictions and decisions (e.g. Kindermans et al., 2019; Ras et al., 2018) – it is exceptionally difficult to anticipate how they will act collectively, including whether their decisions will aggravate or alleviate any devastating effects of collective machine behaviours. What seems clear is that knowing their individual strategies is insufficient for understanding any distinctly collective effects of their interactions in markets – the level of collective machine behaviour is fundamentally different from that of individual strategies. This is what renders the classical sociological vocabulary of collective behaviour relevant, suggesting the possibility that, during collective machine behaviour, even ML algorithms might be overwhelmed and carried away against the goals they pursue. The risk of this happening is likely to be higher in situations in which the ML systems operate dramatically beyond the market situations reflected in the data on which they are trained – resembling a kind of ‘algorithmic anomie’ where known patterns cease to exist, and individual algorithms may be even more prone to follow others in unexpected ways.
Conclusion

As important as it is not to exaggerate new technologies like ML systems, anthropomorphize them or treat them as Frankensteinian monsters, it is also crucial not to under-rate them. In this article, I have argued for seriously considering how ML systems are conceived and designed by market participants who specialize in ML-based securities trading. Their systems come up with trading strategies independently, based on the data they are fed, and the objective function their human designers have equipped them with. Taking this way of operating seriously has important social theory implications. To begin with, it means that in financial markets, trading decisions do not merely originate in algorithmic systems that enact human-defined strategies but also, and increasingly so, in ML systems that develop their own trading strategies, learn from the past and the present and adapt their future orientations accordingly.

Recognizing this entails calls for a sociological discussion of inter-algorithmic relations, or what Rahwan et al. (2019) refer to as collective machine behaviour. This level of collective behaviour demands particular sociological attention, not just because it affects markets and society (as is clear from flash crash events) but also because it paints the contours of a distinct form of ‘machine sociality’, understood as relations playing out between operationally autonomous ML systems whose behaviours can generate emergent collective effects that are irreducible to individual strategies. I have argued that the notion of the social embeddedness of markets – a notion highly relevant for understanding inter-human behaviours – lacks credibility when it comes to collective machine behaviour. By contrast, Goffman’s notion of an interaction order and classical theorization on crowd and collective behaviour are useful heuristics, though more work is needed in this field. That said, the rapidly growing importance of ML systems in society suggests the possibility of a distinct form of machine sociality, urgently calling for a social theory of ML.

Acknowledgements

I thank Kristian Bondo Hansen, Bo Hee Min, Daniel Souleles, Richard Swedberg, Morten Sørensen Thaning, audiences at the Copenhagen Business School and the Stockholm Centre for Organizational Research, as well as three anonymous European Journal of Social Theory reviewers and Gerard Delanty for their valuable comments on earlier versions of this article. I also thank Georgina Kate for her copy-editing assistance.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 725706).
Notes
1. The theoretical discussion presented in this article is based in part on comprehensive empirical work. Since 2014, colleagues (Kristian Bondo Hansen, Pankaj Kumar, Ann-Christina Lange, Bo Hee Min, Nicholas Skar-Gislinge and Daniel Souleles) and I have conducted 213 semi-structured interviews with people who, in various ways, are or have been working with algorithmic trading, either ML-based or not (out of the 213 interviews, I conducted 70 individually or collaboratively). Min and Borch (2021) offer a description of most of these data. However, given the present article’s theory ambitions and the journal’s scope, I deliberately do not mobilize my empirical data explicitly here. Note that the use of ML is not necessarily overlapping with the forms of HFT discussed in existing sociological work (e.g. MacKenzie, 2021). Some firms that specialize in ML would be HFT firms, whereas others would be hedge funds with longer time horizons.

2. By focusing on this term, I also depart from Luhmann since collective behaviour does not play any central role in his work. For a discussion on the status and implications of algorithms – particularly in human–machine interactions – from a Luhmann-inspired point of view, see Esposito (2017).

3. I suspect MacKenzie is led to his conclusion because his analysis primarily attends to human-defined algorithms, indicating that ML-based ones only constitute a small part of what his more than 300 informants specialize in (MacKenzie, 2019, 2021).

References
Abolafia, M. Y. (1996). Making markets: Opportunism and restraint on Wall Street. Harvard University Press.
Aldrich, E. M., Grundfest, J. A., & Laughlin, G. (2017, March 26). The flash crash: A new deconstruction. Retrieved May 15, 2020, from, https://ssrn.com/abstract=2721922
Amoore, L. (2020). Cloud ethics: Algorithms and the attributes or ourselves and others. Duke University Press.
Aradau, C., & Blanke, T. (2017). Politics of prediction: Security and the time/space of governmentality in the age of big data. European Journal of Social Theory, 20(3), 373–391.
Baker, W. E. (1984). The social structure of a national securities market. American Journal of Sociology, 89(4), 775–811.
Berk, R. A. (1974). Collective behavior. Wm. C. Brown.
Blumer, H. (1939). Collective behavior. In R. E. Park (Ed.), An outline of the principles of sociology (pp. 219–280). Barnes & Noble.
Borch, C. (2006). The exclusion of the crowd: The destiny of a sociological figure of the irrational. European Journal of Social Theory, 9(1), 83–102.
Borch, C. (2016). High-frequency trading, algorithmic finance, and the flash crash: Reflections on eventalization. Economy and Society, 45(3–4), 350–378.
Borch, C. (2020). Social avalanche: Crowds, cities and financial markets. Cambridge University Press.
Borch, C., & Lange, A. C. (2017). High-frequency trader subjectivity: Emotional attachment and discipline in an era of algorithms. *Socio-Economic Review*, 15(2), 283–306.

Brayne, S., & Christin, C. (2020). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems: spa004*, 68(3), 608–624.

Brighenti, A. M., & Pavoni, A. (2021). On urban trajectology: Algorithmic mobilities and atmospheric navigation. *Distinktion: Journal of Social Theory*. https://doi.org/10.1080/1600910X.2020.1861044

Buolamwini, J., & Gebru, T. (2018). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. Paper presented at the proceedings of the 1st conference on fairness, accountability and transparency, proceedings of machine learning research.

Coeckelbergh, M. (2020). *AI ethics*. The MIT Press.

Durkheim, E. (1995). *The elementary forms of religious life*. The Free Press.

Edelmann, A., Wolff, T., Montagne, D., & Bail, C. (2020). Computational social science and sociology. *Annual Review of Sociology*, 46(1), 61–81.

Esposito, E. (2017). Artificial communication? The production of contingency by algorithms. *Zeitschrift für Soziologie*, 46(4), 249–265.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin’s Press.

Fourcade, M., & Johns, F. (2020). Loops, ladders and links: The recursivity of social and machine learning. *Theory and Society*, 49(5), 803–832.

Goffman, E. (1963). *Behavior in public places: Notes on the social organization of gatherings*. The Free Press.

Goffman, E. (1983). The interaction order. *American Sociological Review*, 48(1), 1–17.

Golub, A., Keane, J., & Poon, S. H. (2012). High frequency trading and mini flash crashes. Retrieved 20 October, 2014, from http://ssrn.com/abstract=2182097

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3), 481–510.

Guida, T. (2019). *Big data and machine learning in quantitative investment*. John Wiley & Sons.

Hansen, K. B. (2020). The virtue of simplicity: On machine learning models in algorithmic trading. *Big Data & Society*, 7(1), 2053951720926558. https://doi.org/10.1177/2053951720926558

Hansen, K. B. (2021). Model talk: Calculative cultures in quantitative finance. *Science, Technology, & Human Values*, 46(3), 600–627.

Hansen, K. B., & Borch, C. (2021). The absorption and multiplication of uncertainty in machine-learning-driven finance. *British Journal of Sociology*, 72(4), 1015–1029.

Johnson, N., Zhao, G., Hunsader, E., Qi, H., Johnson, N., Meng, J., & Tivnan, B. (2013). Abrupt rise of new machine ecology beyond human response time. *Scientific Reports*, 3(1), 2627.

Kelleher, J. C. (2019). *Deep learning*. The MIT Press.

Kindermans, P. J., Hooker, S., Adebayo, J., Alber, M., Schütt, K. T., Dähne, S., Erhan, D., & Kim, B. (2019). The (un)reliability of saliency methods. In A. Vedaldi, G. Montavon, K. R. Müller, L. Kai Hansen, & W. Samek (Eds.), *Explainable AI: Interpreting, explaining and visualizing deep learning* (pp. 267–280). Springer International Publishing.

Knorr Cetina, K. (2009). The synthetic situation: Interactionism for a global world. *Symbolic Interaction*, 32(1), 61–87.
Knorr Cetina, K., & Bruegger, U. (2002). Traders’ engagement with markets: A postsocial relationship. *Theory, Culture & Society, 19*(5/6), 161–185.

Lang, K., & Lang, G. E. (1961). *Collective dynamics.* Thomas Y. Crowell.

Lange, A. C. (2016). Organizational ignorance: An ethnographic study of high-frequency trading. *Economy and Society, 45*(2), 230–250.

Lange, A. C., Lenglet, M., & Seyfert, R. (2016). Cultures of high-frequency trading: Mapping the landscape of algorithmic developments in contemporary financial markets. *Economy and Society, 45*(2), 149–165.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory.* Oxford University Press.

Le Bon, G. (1960). *The crowd: A study of the popular mind.* The Viking Press.

López de Prado, M. (2018). *Advances in financial machine learning.* Wiley.

Luhmann, N. (1995). *Social systems.* Stanford University Press.

Luhmann, N. (2012). *Theory of society: Volume 1.* Stanford University Press.

MacKenzie, D. (2018a). ‘Making’, ‘taking’ and the material political economy of algorithmic trading. *Economy and Society, 47*(4), 501–523.

MacKenzie, D. (2018b). Material signals: A historical sociology of high-frequency trading. *American Journal of Sociology, 123*(6), 1635–1683.

MacKenzie, D. (2019). How algorithms interact: Goffman’s ‘interaction order’ in automated trading. *Theory, Culture & Society, 36*(2), 39–59.

MacKenzie, D. (2021). *Trading at the speed of light: How ultrafast algorithms are transforming financial markets.* Princeton University Press.

McPhail, C. (1991). *The myth of the madding crowd.* Aldine de Gruyter.

Min, B. H., & Borch, C. (2021). Systemic failures and organizational risk management in algorithmic trading: Normal accidents and high reliability in financial markets. *Social Studies of Science. https://journals.sagepub.com/doi/full/10.1177/03063127211048515*

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism.* New York University Press.

Pardo-Guerra, J. P. (2019). *Automating finance: Infrastructures, engineers, and the making of electronic markets.* Cambridge University Press.

Park, R. E. (1972). *The crowd and the public and other essays.* University of Chicago Press.

Preda, A. (2009). Brief encounters: Calculation and the interaction order of anonymous electronic markets. *Accounting, Organizations and Society, 34*(5), 675–693.

Preda, A. (2017). *Noise: Living and trading in electronic finance.* University of Chicago Press.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C.Pentland, A., ... Wellman, M. (2019). Machine behaviour. *Nature, 568*(7753), 477–486.

Ras, G., van Gerven, M., & Haselager, P. (2018). Explanation methods in deep learning: Users, values, concerns and challenges. In H. J. Escalante, I. Guyon, M. van Gerven, S. Escalera, U. Güçlü, X. Baró, & Y. Güçlütürk (Eds.), *Explainable and interpretable models in computer vision and machine learning* (pp. 19–39). Springer.

Seyfert, R. (2016). Bugs, predations or manipulations? Incompatible epistemic regimes of high-frequency trading. *Economy and Society, 45*(2), 251–277.
Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of go without human knowledge. Nature, 550(7676), 354–359.

Simmel, G. (1950). On the significance of numbers for social life. In K. E. Wolff (Ed.), The sociology of Georg Simmel (pp. 87–104). The Free Press.

Simmel, G. (1992). Soziologie. Suhrkamp.

Simmel, G. (1997). The metropolis and mental life. In D. Frisby & M. Featherstone (Eds.), Simmel on culture: Selected writings (pp. 174–185). SAGE.

Smelser, N. J. (1962). Theory of collective behavior. The Free Press.

Sornette, D., & von der Becke, S. (2011). Crashes and high frequency trading: An evaluation of risks posed by high-speed algorithmic trading. Swiss Finance Institute, Research Paper Series No. 11–63.

Souleles, D. (2019). The distribution of ignorance on financial markets. Economy and Society, 48(4), 510–531.

Svetlova, E. (2021). AI meets narrative: The state and future of research on expectation formation in economics and sociology. Socio-Economic Review. https://doi.org/10.1093/ser/mwab033

Tarde, G. (1968). Penal philosophy. Patterson Smith.

Tarde, G. (1989). L’opinion et la foule. Presses Universitaires de France.

Tilly, C. (1978). From mobilization to revolution. Addison-Wesley Publishing Company.

Tilly, C. (2004). Social movements, 1768–2004. Paradigm Publishers.

Turner, R. H., & Killian, L. M. (1957). Collective behavior. Prentice-Hall.

Wagner-Pacifici, R., Mohr, J. W., & Breiger, R. L. (2015). Ontologies, methodologies, and new uses of big data in the social and cultural sciences. Big Data & Society, 2(2). https://doi.org/10.1177/2053951715613810

Author biography

Christian Borch is a Professor of Economic Sociology and Social Theory at the Copenhagen Business School, Denmark. His research focuses on algorithmic trading, machine learning, crowd theory and social theory. His most recent book is Social Avalanche: Crowds, Cities and Financial Markets (Cambridge University Press, 2020).