CMA
Competition & Markets Authority

Algorithms:

How they can reduce competition and harm consumers

© Crown copyright 2021

You may reuse this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence.

To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/ or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

Contents

Page

Executive Summary

We spend much of our lives online, be it consuming news, socialising, dating, ordering food, or arranging travel. Many of these online activities and the markets that underpin them could not exist without algorithms, often in the form of artificial intelligence. Algorithms have enabled considerable gains in efficiency and effectiveness, such as repricing portfolios of thousands of products in real time. Importantly, algorithms are at the heart of many technology companies, including some of the world's most strategically significant firms. However, algorithms can be used in ways that reduce competition and harm consumers. As algorithmic systems become more sophisticated, they are often less transparent, and it is more challenging to identify when they cause harm.

The publication of this paper, and the accompanying call for information, mark the launch of a new CMA programme of work on analysing algorithms, which aims to develop our knowledge and help us better identify and address harms. This paper reviews the potential harms to competition and consumers from the use of algorithms, focussing on those the CMA or other national competition or consumer authorities may be best placed to address.

We first describe direct harms to consumers, many of which involve personalisation. Personalisation can be harmful because it is difficult to detect either by consumers or others, targets vulnerable consumers or has unfair distributive effects. These harms often occur through the manipulation of consumer choices, without the awareness of the consumer.

The paper then explores how the use of algorithms can exclude competitors and so reduce competition (for example, a platform preferencing its own products). We outline the most recent developments in the algorithmic collusion literature; collusion appears an increasingly significant risk if the use of more complex pricing algorithms becomes widespread. We also describe how using ineffective algorithms to oversee platform activity fails to prevent harm.

Next, we summarise techniques that could be used to analyse algorithmic systems. Potentially problematic systems can be identified even without access to underlying algorithms and data. However, to understand fully how an algorithmic system works and whether consumer or competition law is being breached, regulators need appropriate methods to audit the system. We finally discuss the role of regulators. Regulators can help to set standards and facilitate better accountability of algorithmic systems, including support for the development of ethical approaches, guidelines, tools and principles. They can also use their information gathering powers to identify and remedy harms on either a case-by-case basis or as part of an *ex-ante* regime

overseen by a regulator technology firms, such as the proposed Digital Markets Unit (DMU) in the UK.

More research, as ever, is needed to assess and quantify these areas of harm because digital markets are evolving rapidly. Some harms may be particularly challenging to identify. For example, it is difficult to see the combined effect of smaller non-price ‘nudges’ such as presenting expensive products first to some consumers, which may have the same effect as the personalisation of listed prices, but be harder to detect. Even in relatively well-researched areas, such as algorithmic collusion, there is a dearth of empirical studies to understand real-world impacts.

1. Introduction

1.1 Algorithms are sequences of instructions to perform a computation or solve a problem. We use the term 'algorithm' to include simpler sets of rules as well as more advanced machine learning or artificial intelligence (AI) code. In this paper, we use a broad interpretation of the term 'algorithmic system', as a convenient shorthand to refer more widely to automated systems, a larger intersection of the algorithm, data, models, processes, objectives, and how people interact and use these systems.[1]

1.2 Simple algorithmic systems are not new. They have been used for decades to efficiently manage business processes and automate simple decision-making processes – particularly in manufacturing, supply chain logistics, and making pricing decisions.

1.3 The widespread availability of increasingly large volumes of granular data, combined with the increasing computing power to process it, has meant that more complex processes can also be automated. Machine learning and AI are now employed in a wide range of contexts, industries and applications.[2] Algorithms are at the heart of some of the largest and most strategically significant firms' operations, for example Google's search algorithm and Facebook's News Feed algorithm. Even small businesses are increasingly using machine learning by buying tools developed by third parties.[3] For example, the Amazon Web Service Marketplace contains hundreds of machine learning services in areas such as speech recognition, automatic facial recognition, document summarisation and many more. These trends seem likely to continue, as businesses both big and small make use of better technologies to enable product innovation and improve their internal processes.[4]

1.4 Many algorithmic systems provide substantial benefits to consumers. People now spend much of their lives online, be it consuming news, socialising, dating, ordering food, or arranging travel. Algorithmic systems can provide individualised recommendations that are relevant, saving people time, and allowing them to focus more on what matters to them. Many of the products that enable these activities could not exist without algorithms and the data that powers them. Businesses can use algorithmic systems to

---

[1] World Wide Web Foundation (2017), 'Algorithmic Accountability – Applying the concept to different country contexts'.
[2] See, for instance, McKinsey Global Institute (2018), 'Notes from the AI Frontier – Insights from Hundreds of Use Cases', *Discussion paper*.
[3] The 2018 State of Small Business Britain report found that 9 percent of micro-businesses (1-9 employees) were using machine learning. This was an increase from 3 percent in 2012 and is likely to grow further. See Roper, S and Hart, M (2018), 'The State of Small Business Britain Report 2018'. Enterprise Research Centre.
[4] El-Hanfy, S and Webster, Z (2020), 'New business horizons in the AI landscape'. Innovate UK.

optimise their actions and interfaces with customers (often referred to as 'choice architecture', i.e. the process and outcome of design decisions about how choices are presented to people (including for user interfaces in an online environment), and the impact of that presentation on people's decisions). Such optimisation can be beneficial, as it may enhance the quality of products and services for consumers and allow the company to make effective improvements based on empirical evidence.

1.5 Algorithmic systems also benefit consumers indirectly by increasing efficiency and effectiveness across many areas. For example, businesses can now use algorithmic systems to reprice portfolios of thousands of products in real-time[5] and pricing efficiencies can be passed on to customers. As third-party solutions make algorithmic pricing more accessible to smaller businesses, this can lower the barriers to entry, and enhance innovation and competition. Businesses can also use algorithmic systems to mitigate online harms[6] and prevent discrimination against particular groups[7] to provide consumers with better quality (and less harmful) products. Algorithmic systems can even be used to detect collusion between firms to ensure competitive prices.[8]

1.6 However, algorithms can also cause harm and this is the focus of this paper. As a preliminary remark, we note that some algorithmic systems are complex, especially those involving machine learning algorithms, and their behaviour and harms may not be perfectly anticipated by developers and firms. Nevertheless, firms are responsible for effective oversight of such systems, which should include robust governance, holistic impact assessments, monitoring and evaluation.

1.7 The CMA's mission is to make markets work well for consumers, business, and the economy. The CMA works to promote competition for the benefit of consumers, to enable consumers to get a good deal when buying goods and services and to ensure that businesses operate within the law. The CMA also has a responsibility to investigate mergers between organisations to make sure they don't reduce competition, to investigate entire markets if it thinks there are competition or consumer problems, to take action against businesses that take part in cartels or anti-competitive behaviour, and to protect consumers from unfair trading practices.

---

[5] Competition and Markets Authority (2018), 'Pricing algorithms: Economic working paper on the use of algorithms to facilitate collusion and personalised pricing'.
[6] Cambridge Consultants (2019), 'Use of AI in Online Content Moderation'. *Ofcom*.
[7] Kleinberg, J, Ludwig, J, Mullainathan, S, & Sunstein, CR (2020), 'Algorithms as discrimination detectors' in *Proceedings of the National Academy of Sciences*.
[8] Johnson, J & Sokol, DD (2020), 'Understanding AI Collusion and Compliance' in *Cambridge Handbook of Compliance, (D. Daniel Sokol & Benjamin van Rooij, editors), (Forthcoming)*.

1.8 A new Digital Markets Unit (DMU) is being set up within the CMA to implement a pro-competitive regime for digital markets. (See the Government response to the CMA digital advertising market study.) In its advice to government on what functions and powers the DMU will need, the CMA Digital Markets Taskforce recommended an *ex-ante* regime. This would proactively prevent harms on an ongoing basis for the most powerful digital firms and gives the Unit a proactive monitoring role in the wider digital market. As part of this, it will be important to monitor developments in the application of machine learning and AI to ensure they do not lead to anti-competitive behaviour or consumer detriment, particularly in relation to vulnerable consumers.[9] The CMA already has experience in this area[10] and an increasing proportion of our work in competition policy, consumer protection, and regulatory design will involve understanding the operation and effects of key algorithms and automated decision systems of significant firms.

1.9 Section 2 of this paper gives an overview of some of the potential harms to competition and consumers that may arise from misuse of algorithmic systems, focusing on the perspectives of economics, computer science, and behavioural science. (The paper does not focus on which specific parts of UK consumer and competition legislation could be used to address each of these potential harms.) It builds on our 2018 economic working paper on the use of algorithms to facilitate collusion and personalised pricing, but goes beyond that paper to discuss other aspects of personalisation, discrimination, exclusionary practices, and harms arising from ineffective oversight by platforms. Many of the harms discussed in this paper are not new, but recent advances in technology mean that many algorithms today can enact changes at a scale that makes these harms more pronounced.

1.10 Section 3 of this paper discusses some of the techniques that could be used to investigate the harms, and Section 4 discusses the potential role of regulators in addressing them.

1.11 The issues in this paper are our current view of the most significant harms to consumers and competition arising from algorithmic systems, and what should be done about them. We intend to present the key issues, stimulate debate and encourage the gathering of intelligence from academia, the competition community, firms, civil society and third sector organisations. As

---

[9] See Strategic Recommendation D in 'Unlocking digital competition, Report of the Digital Competition Expert Panel', March 2019 (available here) and also the CMA's Digital Markets Strategy.
[10] See, for example, the CMA's Economic working paper on the use of algorithms to facilitate collusion and personalised pricing and the CMA investigation into the use of automated repricing software to facilitate an illegal cartel.

we develop our approach to investigating, preventing and remedying these harms, we invite views on:

(a) Whether these areas and potential harms are the right ones to focus on, their likelihood and impact, and whether there are others that deserve attention;

(b) Methods and techniques to i) monitor and assess the potential for harms and ii) investigate and audit the algorithms and systems of any given firm; and

(c) Feasible, effective and proportionate measures to prevent or remedy algorithmic harms.

1.12 In deciding where to focus our efforts and resources, we will apply our prioritisation principles. More specifically, in the context of which algorithmic harms or issues to prioritise, we note that (all else equal) the risk and impact of harm will be greater where a single algorithm or model is applied on a large scale, either because it is an 'industry standard' used by most market participants or because it is used by a firm with significant market power. The harm is also likely to be greater where algorithmic systems inform decisions that could have significant effects on consumers (such as decisions about jobs, credit, and housing).

2. Theories of Harm

2.1 This section sets out various 'theories of harm' regarding how the use of algorithms and automated decision-making systems could give rise to anti-competitive effects or consumer detriment.

2.2 The potential harms from the use of algorithms are wide ranging and raise issues in multiple overlapping areas of public policy and regulation (for example, in criminal justice, public health and 'online harms'). We focus primarily on economic harms that could be addressed by enforcement of competition and consumer law, or via new powers of the DMU. However, we also note that the use of algorithmic systems can cause harms to people in their role as citizens, and not just their role as consumers or business owners (for example, in election integrity or media plurality), but we do not discuss these wider harms in this paper. We also do not discuss in detail the legal tools that could be used to tackle the different harms. We note that several of these harms overlap and may be tackled through a combination of approaches and legal tools, or in collaboration with other regulators and use of their expertise and powers. These include the Information Commissioner's

Office (ICO), Financial Conduct Authority (FCA), Ofcom, and the Equalities and Human Rights Commission (EHRC).

2.3 Many of the harms discussed in this section involve the use of algorithmic systems to optimise businesses' actions and interfaces with customers, often referred to as choice architecture. For example, the position of the "Buy" button on a shopping website, the colour of an information banner and a default payment method would all be examples of choice architecture based on algorithms. Although choice architecture can be used to benefit consumers, firms may instead exploit inherent weaknesses in consumers' ability to engage in markets in ways which go against consumers' interests and undermine competition[11] (for example, by exploiting their limited attention, loss aversion, or inertia, leading to susceptibility to default options).[12] Businesses can use increasingly sophisticated techniques to understand and exploit these biases[13] in ways that harm consumers directly (such as making purchasing decisions that they would not under different choice architecture)[14] or better enable incumbent firms to exclude or marginalise competitors (for example if they engage in non-transparent self-preferring behaviour). These harmful user interface design choices are known as 'dark patterns'.

2.4 Many of the harms we discuss also involve personalisation. The availability of ever-greater volumes of data about consumers, coupled with the use of algorithmic systems, has resulted in the ability of firms to personalise their actions and interfaces to each consumer to an extent not previously possible. As firms analyse the characteristics and behaviour of consumers, they can use machine learning techniques such as clustering algorithms to identify meaningful categories of consumers, classify new and existing customers according to these categories, and apply different treatments to each category.[15] Such personalisation also extends to the choice architecture. Examples of this include which options are presented, in what order, how

---

[11] In many cases where firms design choice architecture that interacts with consumers' behavioural biases, competition may not result in good outcomes for consumers, as firms that engage in such practices may derive unfair advantages that enable them to out-compete and 'drive out' firms that decline to do so. In this way, markets can converge on 'bad equilibria', and these market failures can justify regulatory intervention.
[12] It is well-established that consumers are subject to behavioural biases and that firms can exploit these biases for profit. See, for instance, Walker, M (2017), 'Behavioural economics: the lessons for regulators', *European Competition Journal*, *13*(1), 1-27. See also Spiegler, R (2011), *Bounded rationality and industrial organization*. Oxford University Press.
[13] For a detailed example of some of the approaches used by firms in digital sector to exploit consumers' biases, see the discussion of the CMA's Hotel Online Booking case in Fung, S, Haydock, J, Moore, A, Rutt, J, Ryan, R, Walker, M, & Windle, I (2019), 'Recent Developments at the CMA: 2018–2019', *Review of Industrial Organization*, *55*(4), 579-605. A common thread across the practices in that case was that they were misleading to consumers.
[14] Calo, R (2013), 'Digital market manipulation', *Geo. Wash. L. Rev.*, *82*, 995.
[15] For instance, firms could use machine learning to estimate heterogenous treatment effects, and test which customer or user groups will engage optimally with a product or service. See Wager, S and Athey, S (2018), 'Estimation and Inference of Heterogeneous Treatment Effects using Random Forests', *Journal of the American Statistical Association* 113(523). In addition, see the literature that proceeds the paper.

prices are presented, the timing and content of notifications and reminders, and indeed whether and which offers are made.[16] Consumers may have limited visibility or knowledge of the inferences the firms may draw about them and the categories in which they are placed. In addition, whilst personalisation can be beneficial for consumers, for example by allowing firms to provide services and information that are tailored and relevant to each individual, it can also lead to harmful and unacceptable discrimination. This is particularly the case if firms use (unwittingly or otherwise) categories that are correlated with consumer vulnerability[17] and protected characteristics.[18]

2.5 Although most of the harms discussed in this paper relate to consumer-facing systems, some relate to exclusionary practices by dominant firms that can exclude or marginalise competitors, and indirectly harm consumers.

2.6 Our discussion of potential harms is organised as follows:

(a) We first discuss direct harms to consumers from algorithmic systems, including:

(i) how algorithmic systems can be used to personalise prices in a way that is opaque to the consumer;

(ii) more general personalisation, where algorithmic systems can be used to manipulate choice architecture or customer journeys more widely;

(iii) algorithmic discrimination, a harmful form of personalisation regarding protected characteristics, and what we can learn from the burgeoning field of algorithmic fairness; and

(iv) unfair ranking and design, including how algorithmic systems can be used to facilitate the preferencing of others for commercial advantage.

(b) We then discuss the use of algorithmic systems in exclusionary practices, where dominant firms can take actions that deter competitors from challenging their market position. Such practices include self-preferencing, manipulating ranking algorithms to exclude competitors, and

---

[16] A detailed discussion of choice architecture with respect to digital advertising, including definitions and examples, can be found in Appendix Y of the CMA's Market Study into Digital Advertising.
[17] The Money and Mental Health Policy Institute has recently researched how the frictionless, pressuring and personalised design of online shopping sites can make it challenging for people with mental health problems to control spending. (Money and Mental Health Policy Institute (2020), 'Convenience at a cost'.)
[18] Furthermore, it is well understood that where firms can personalise any aspect of their offering to its customers, firms can focus their efforts to compete for active, engaged, marginal consumers and extract profits from inert, passive, infra-marginal consumers. In this way, the protection afforded to consumers by the disciplining effect of competition due to the vigilance of marginal consumers is eroded for inert consumers. See Chisolm, A (2016), 'Consumer engagement in a digital world'.

changing an algorithmic system in a gateway service that unintentionally harms business that rely on it.

(c) We then briefly discuss the issue of potential collusion by pricing algorithms.

(d) We end with a discussion of ineffective platform oversight, where a lack of transparency can make it difficult to externally evaluate whether an algorithmic system is effective, and therefore drive improvements.

2.1 Direct harms to consumers

2.7 In this section, we consider direct harms to consumers that are related to unfairness, manipulation and exploitation. Algorithmic systems can be used to personalise a consumer's experience of a service. If consumers are not aware that it is occurring, or if it gives rise to unfair distributive effects or harms consumers who are vulnerable, it is more likely to be exploitative. In addition, there are potential consumer harms that are not related to personalisation, but instead relate to unfair ranking and design of online services.

2.8 In the UK, traders have a general duty not to trade unfairly by acting contrary to the requirements of professional diligence which distorts the average consumer's decisions in relation to a product or service. This can be broadly understood as failing to act in accordance with acceptable trading practice that a reasonable person would expect. In addition, misleading and aggressive practices are prohibited. This includes omission of material information from consumers which impairs their ability to make an informed choice.

2.1.1 Personalised pricing harms

2.9 Personalised pricing includes advertising different prices to different people and practices which achieve the same effect, such as providing discounts to selected customers. Firms personalise prices based on what it thinks different customers are willing to pay, in order to increase profits.

2.10 One of the key characteristics of digital markets is the ability to access vast volumes of personal data and apply analytics to approximate the user's willingness to pay. Such practices may be particularly powerful when digital market providers, for example platforms, control the interface that the consumer interacts with, and can limit the alternative options a consumer can access and therefore limit the ability to switch to another provider.

2.11 In many cases, personalised pricing can be beneficial, increasing total output and consumer welfare. For example, personalised pricing can lower search costs for consumers and bring about a more precise match between consumers and products and services. It may also allow firms to set a lower price and profitably sell to consumers that would not be willing to pay the uniform price that firms would otherwise set. Similarly, the ability to offer targeted discounts might help new entrants to compete, particularly in markets with switching costs.

2.12 However, there are other situations where personalised pricing could lead to consumer harm. The conditions under which competition authorities might be concerned about personalised pricing are outlined in an OFT economics paper in 2013, and include where there is insufficient competition (i.e. monopolist price discrimination), where personalised pricing is particularly complex or lacking transparency to consumers and/or where it is very costly for firms to implement.[19] In addition, personalised pricing could harm overall economic efficiency if it causes consumers to lose trust in online markets. It could also be harmful for economic efficiency when personalised pricing increases search and transaction costs, such as consumers needing to shop around or take significant or costly steps to avoid being charged a premium.[20]

2.1.1.1 Empirical evidence of personalised pricing

2.13 Personalised advertised prices are an overt form of personalised pricing. There is currently limited evidence of their use by online retailers. According to a summary provided by the European Commission to an OECD committee, no personalised advertised pricing has been found in the EU on any significant scale.[21] Citizens Advice (August 2018) found that personalised advertised pricing was not widespread in essential markets (energy, water, telecoms and post), but that this could change quickly.[22]

2.14 Empirically, there has also been limited consistent review showing how personalised advertised prices are used. Some examples exist, such as

---

[19] Office of Fair Trading (2013), 'The economics of online personalised pricing'. The CMA's position on personalised pricing is also set out in our contribution to an OECD discussion. See 'Personalised Pricing in the Digital Era – Note by the United Kingdom'.
[20] Borgesius, FZ, & Poort, J (2017), 'Online price discrimination and EU data privacy law', *Journal of consumer policy*, *40*(3), 347-366.
[21] Personalised Pricing in the Digital Era – Note by the European Union, paragraphs 27 to 36 provide a good summary of studies, surveys, and mystery shopping exercises carried out by EU competition and consumer protection authorities between 2013 to 2018. The most recent and largest of these, carried out on behalf of the European Commission's Directorate-General for Justice and Consumers, found no evidence of consistent and systematic online personalised pricing across eight EU member states, including the UK. (Ipsos, London Economics & Deloitte, report for DG JUST (2018), 'Consumer market study on online market segmentation through personalised pricing/offers in the European Union', 19 July.)
[22] Citizens Advice (2018), 'A price of one's own – an investigation into personalised pricing in essential markets'.

where B&Q, a British home improvement company, employed dynamic pricing through digital price tags in its shops. These price tags used information from customers' phones to adjust the displayed price based on the customer's spending habits and loyalty card data.[23] Several older papers have also found some evidence of price discrimination occurring in a limited way (e.g. Hannák et al. 2014; Mikians et al 2013, Mikians et al. 2012).[24]

2.15 If personalised advertised pricing is as limited as it appears to be, this may be due to businesses' concerns about the potential reputational impact of personalised pricing and reflect a belief that consumers will view personalised pricing as unfair.[25] They may therefore employ other techniques to personalise prices that are harder for consumers to detect.

2.16 For instance, the use of loyalty programs and promotional offers (such as coupons and vouchers) to offer personalised effective prices is common and well-established, particularly in retail and many other business-to-consumer markets. However, this can have a downside for less active consumers. Examples include firms charging higher prices to longstanding customers than new customers or those who renegotiate their deal,[26] and 'price walking', the practice of increasing prices to consumers each year at renewal. Usually, differential prices are framed in terms of discounts rather than surcharges, which may make them more acceptable to consumers. Such practices are also relatively well-known, and many consumers may take advantage of lower prices by switching often.[27]

2.1.1.2 Complex and opaque pricing techniques

2.17 Firms can use a wide range of data about consumers to support inferences about their willingness-to-pay and personalise prices accordingly. This data

---

[23] Thomas, S (2014) 'Does dynamic pricing risk turning personalisation into discrimination?' 22 October.
[24] Hannák, A, Soeller, G, Lazer, D, Mislove, A, & Wilson, C (2014), 'Measuring Price Discrimination and Steering on E-commerce Web Sites', in *Proceedings of the 2014 conference on internet measurement conference* (pp. 305-318). Mikians, J, Gyarmati, L, Erramilli, V, & Laoutaris, N (2013), 'Crowd-assisted Search for Price Discrimination in E-Commerce: First results', in *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies* (pp. 1-6). Mikians, J, Gyarmati, L, Erramilli, V, & Laoutaris, N (2012), 'Detecting price and search discrimination on the Internet', in *Proceedings of the 11th ACM workshop on hot topics in networks* (pp. 79-84).
[25] Office of Fair Trading (2013), 'Personalised Pricing – Increasing Transparency to Improve Trust', p.15. Office of Fair Trading (2010), 'Online Targeting of Advertising and Prices – A market study', paragraphs 6.11 to 6.14. See also Obama White House (2015), 'Big Data and Differential Pricing', p.13.
[26] In 2018, the CMA responded to a super-complaint by Citizens Advice raising concerns about long term customers paying more for goods and services. CMA (2018), 'Tackling the loyalty penalty', 19 December.
[27] Ultimately, consumers who pay the higher advertised price are also receiving a price that is 'personalised' to their behaviour or characteristics (i.e. not taking actions or having the features that would qualify them for discounts). Judging by popular dissatisfaction with 'loyalty penalties', some consumers find higher relative prices through inaction to also be objectionable.

may be collected and used by firms in ways which consumers do not expect or have little control over.

(a) For example, multiple press reports have speculated on whether ridesharing services like Uber personalise prices based on factors such as payment method.[28] In 2017, it was reported that Uber's 'route-based pricing' charges customers based on what it predicts they are willing to pay,[29] and that Uber attempts to sort customers into 'time-sensitive' and 'price-sensitive' riders.[30] In 2016, Uber's head of economic research stated that Uber found people are more likely to pay higher 'surge' prices if their phone is almost out of battery.[31]

(b) MGM casinos in Las Vegas were found to use customer data to personalise promotions based on their profitability. MGM used loyalty cards and linked casino playcards to capture where, when, how long and how much customers were playing, in addition to other activities in the casino.[32] According to Stucke and Ezrachi (2020), the data helped casinos to offer customers the cheapest mix of perks to incentivise customers to spend the most money over the customer's lifetime.

(a) It has been alleged that Staples' website displayed different prices to people, depending on how close they were to a rival brick-and-mortar store belonging to OfficeMax or Office Depot.[33,34] In general, we expect firms to respond to competitive conditions applicable to each of their customers when setting personalised prices. All else equal, if some customers have more options than others, because there are more available firms that compete for them, those customers receive a lower personalised price.

2.18 Firms can also use machine learning and data science techniques to reduce customer attrition (or 'churn'), by analysing what characteristics or behaviours of their customers are predictive of exit or switching.[35] These churn models may then inform firms' decisions about whether and how much to increase prices, what price to offer on renegotiation, and 'win-back' offers. Aside from

---

[28] The Guardian (2018), 'Is your friend getting a cheaper Uber fare than you are?', 13 April.
[29] Bloomberg (2017), 'Uber Starts Charging What It Thinks You're Willing to Pay', 19 May.
[30] Business Insider (2017), 'Uber may charge you more based on where you're going', 20 May.
[31] NPR (2016), 'This Is Your Brain On Uber', 17 May.
[32] Stucke, ME & Ezrachi, A (2020), 'Competition Overdose: How Free Market Mythology Transformed Us from Citizen Kings to Market Servants'.
[33] Wall Street Journal (2012), 'Websites Vary Prices, Deals Based on Users' Information', 24 December.
[34] These online pricing practices would be consistent with the findings of the US FTC, in their successful challenge to the proposed merger of Staples and Office Depot in 1997, that prices in physical stores also depended on whether there was a rival store in the local area. (Baker, JB (1999), 'Econometric analysis in FTC v. Staples', *Journal of Public Policy & Marketing*, *18*(1), 11-21.)
[35] Chen, Z-Y, Fan, Z-P & Sun, M (2012), 'A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioural data'. *European Journal of Operational Research* 223(2).

any concerns about the lack of transparency of these practices for consumers, they can also result in distributional outcomes that are viewed as unacceptable, particularly if those who pay more are likely to be 'vulnerable'.[36] (These practices can also have implications for competition and the ability of incumbents to exclude rivals, which are discussed in section 2.2.3 on predatory pricing below.)

2.19 For instance, the FCA found that some home and motor insurance firms use complex and opaque pricing techniques to identify consumers who are more likely to renew with them. These firms then increase prices to these customers at renewal each year, resulting in some consumers paying very high prices, many of whom are unaware of this. These practices disproportionately affect older consumers and cumulatively over time can lead to large increases in price, of the order of 100 percent or more.[37]

2.20 It is also feasible that personalisation could be applied to other types of pricing choice architectures. Examples include drip pricing, where prices are displayed without fees or additional charges, which is unlawful under consumer protection law, and reference pricing, where the price of a product is displayed in contrast to other more expensive bundles of the same product.

2.1.2 Harms from non-price personalisation

2.21 In this section, we discuss a number of different ways that choice architecture and other non-price information may be personalised. If consumers are not aware of how or that it is occurring, or if it gives rise to unfair distributive effects or harms consumers who are vulnerable, it is more likely to be problematic.

2.1.2.1 Personalised rankings

2.22 It is commonplace for firms to present consumers with a set of options or results that are relevant to them. For example, ecommerce websites can present search results for products in response to a search query. Firms may use information about the user beyond the search query to decide which results to display and in what order. This information could include the user's location, their previous queries, and their previous browsing and purchase behaviour.

---

[36] A specific example of this is the "loyalty penalty", where longstanding customers were paying more than new customers. See CMA's ongoing work into loyalty penalty complaints, in collaboration with other regulatory bodies.
[37] Financial Conduct Authority (2020), 'General insurance pricing practices market study: Final Report'.

2.23 Consumers are likely to be affected by position bias and cannot always observe exactly how or why search results are presented in any given order. Firms could therefore manipulate consumers into making different decisions that are more profitable for the firm, but which the consumer would not have made under more objective or neutral conditions. It could also lead to 'price steering', which could achieve (albeit indirectly) similar results as personalised pricing, by presenting higher priced products to consumers with a higher willingness-to-pay.

2.24 In contrast to personalised advertised prices, which appears to be uncommon, personalised rankings and results are widespread. For example, a study conducted by the European Commission in 2018 found that 61 percent of the 160 e-commerce sites it visited in a mystery shopping exercise personalised ranking of search results.[38]

2.25 Rankings can be unfair in other ways, without being personalised. These are discussed in a later section on 'Unfair ranking and design'.

2.1.2.2 Recommendation and filtering algorithms

2.26 Recommendation and collaborative filtering algorithms are a general class of systems that affect choice architecture by determining the range of information and options presented to a consumer. Examples include those used by social media platforms and video/audio streaming platforms to determine which content to show or suggest to users, and also by ecommerce firms to suggest relevant products to consumers.

2.27 Although not a focus of this paper, we note that there have been significant concerns[39] with recommendation and collaborative filtering systems including: that they may be exacerbating self-control problems and lead to overuse by consumers;[40] that they may promote harmful (but engaging) content where platforms have not filtered them out;[41] and that they may limit the range of options that users are exposed to and lead to fragmentation of shared public discourse and understanding of reality ('echo chambers' and 'filter bubbles').[42]

---

[38] European Commission, 'Consumer market study on online market segmentation through personalised pricing/offers in the European Union', 19 July 2018, available here.
[39] For a recent review, see Milano, S, Taddeo, M, & Floridi, L (2020), 'Recommender systems and their ethical challenges', *AI & SOCIETY*, 1-11.
[40] Hasan, MR, Jha, AK, & Liu, Y (2018), 'Excessive use of online video streaming services: Impact of recommender system use, psychological factors, and motives', *Computers in Human Behavior*, *80*, 220-228.
[41] For a typical example of these kinds of concerns raised in the press, see Wired (2018), 'Up Next: A Better Recommendation System', 4 November. See also analysis conducted by the Guardian and Guillaume Chaslot which scraped YouTube recommendations (The Guardian (2018), 'How an ex-YouTube insider investigated its secret algorithm', 2 February).
[42] Vestager, M (2020), 'Algorithms and democracy – AlgorithmWatch Online Policy Dialogue', 30 October.

2.28 For example, there is some evidence to suggest that addictive technologies such as social media have harm-producing characteristics, which pose a challenge for antitrust enforcement.[43] Rosenquist et al. (2020) argue for a reframing of assumptions around consumer welfare away from the idea that increased consumption leads to increased utility. The negative impacts of addictive technologies related to increased consumption of social media,[44] for example, effectively lower the quality of the product or service, which lowers consumer welfare.

2.29 Additionally, the autosuggestion feature of search engines such as Google and Bing can be susceptible to manipulation through adversarial attacks, given that they depend on users' past queries.[45] Coordinating a large number of specific queries can make it possible for an entity to manipulate the list of suggestions, which can act as signals to consumers and affect their perception of the market; not only for goods and services but also for ideas.

2.1.2.3 Manipulating user journeys

2.30 Choice architecture also includes whether and when to send notifications and prompts, and other aspects of the user experience and journey. In addition, choice architecture may be personalised and lead to distortions that undermine consumer choice and competition.

2.31 For example, firms use a variety of techniques to predict the likely rating that a user would give of their service, based on analysing their users' characteristics and usage/interaction history. Based on this information, firms can manipulate and bias the group of consumers that contribute a rating for the product or service, by personalising whether and when they receive notifications and prompts to leave a rating. For example, companies can target particular consumers at a particular time at which they are more likely to give them positive reviews on an app.[46] This practice has led to ratings inflation, and less useful and informative ratings for consumers generally.

---

[43] Rosenquist, JN, Scott Morton, FM & Weinstein, SN (2020), 'Addictive Technology and its Implications for Antitrust Enforcement'. *Yale School of Management*. September.
[44] Such as purchasing products or watching content that consumers later regret or causes them harm, leading to negative emotions. See Rosenquist et al. (2020), 'Addictive Technology and its Implications for Antitrust Enforcement'.
[45] Hazen, TJ, Olteanu, A, Kazai, G, Diaz, F & Golebiewski, M (2020), 'On the Social and Technical Challenges of Web Search Autosuggestion Moderation'. *Preprint*. 13 July.
[46] McGee, P. (2020) 'Apple: how app developers manipulate your mood to boost ranking'. *The Financial Times*. Available here.

2.1.3 Algorithmic discrimination

2.32 Misuse of algorithmic systems to personalise services and offers to consumers can result in illegal discrimination. In the UK, equality law generally prohibits people and organisations providing services from discriminating on the basis of 'protected characteristics'.[47,48] This includes indirect discrimination, which are situations where a policy applied equally nevertheless has disproportionately harmful effects on a particular group without objective justification (i.e. a proportionate means of achieving a legitimate aim).[49]

2.33 Equality law, in particular through the concept of indirect discrimination, prohibits many discriminatory effects of algorithmic systems that have been reported in recent years.[50] However, enforcement is difficult,[51,52] and to date, no legislation has been passed in the UK that has been designed specifically to tackle discrimination in AI systems.[53] Nonetheless, data protection law, through the General Data Protection Regulation (GDPR), contains the fairness principle, which requires the ICO to intercede on issues where an individual is unjustly affected. Therefore, the processing of personal data that leads to unjust discrimination would contravene the fairness principle under the GDPR.

2.34 As discussed, in the UK, traders have a general duty not to trade unfairly and act in accordance with the requirements of professional diligence. Compliance with equality and data protection law is part of these requirements.

2.35 It should be noted that the increasing use of algorithmic decision-making holds great promise for enhancing fairness. Algorithms can be subject to greater and more in-depth scrutiny than human decision-makers, and it can

---

[47] Equality and Human Rights Commission, 'Delivering services and the law'.
[48] There are some general exceptions, including services for particular groups, as well as industry-specific exceptions (for instance, age discrimination is permitted in financial services). See Equality and Human Rights Commission, 'Equality law – Banks and other financial services providers'.
[49] Equality and Human Rights Commission, 'What is direct and indirect discrimination?'.
[50] Zuiderveen Borgesius, F (2018), 'Discrimination, artificial intelligence, and algorithmic decision-making', Council of Europe.
[51] *ibid*. Specifically, Zuiderveen Borgersius highlights that the prohibition of indirect discrimination does not provide a clear and easily applicable rule. It can be difficult to demonstrate that a seemingly neutral policy disproportionately affects a protected group, particularly where it is infeasible for consumers or researchers to access or compile data about outcomes across large samples of individuals. Also, whether an alleged discriminator has an 'objective justification' often depends on a nuanced proportionality test that takes account of all the context of a case.
[52] In addition, UK equality law does not have extraterritorial jurisdiction, unlike EU competition law and data protection law after GDPR. An Information Society Service Provider (ISSP), i.e. someone providing services through a website such as online shopping or advertising, that is not based in the UK would not be subject to UK equality law. (See Equality and Human Rights Commission, 'Advertising and marketing', and id., 'Websites and Internet services'.)
[53] Allen, R (2020), 'Artificial Intelligence, Machine Learning, algorithms and discrimination law: the new frontier', paper prepared for Discrimination Law in 2020 conference, Congress House, 31 January 2020.

be easier to measure, amend and correct biases in algorithmic systems than in humans.[54]

2.36 There is a large, interdisciplinary literature studying algorithmic discrimination, strategies for detection and mitigation,[55] and explaining automated decisions.[56] A thorough account of the main questions, issues and contributions in this literature is beyond the scope of this paper. However, we follow this literature with interest, because:

(a) The techniques and approaches used to collect data and analyse algorithmic systems for discrimination can be generalised or also applied to analyse other social and economic effects, including some of the other theories of harm that we have discussed. We explore these further in section 3 of this paper.

(b) The application of consumer law to protect people against discrimination arising from misuse of algorithmic systems is a relatively unexplored area, compared with the role of anti-discrimination and data protection law.[57] Discrimination can harm some consumers, including in economic contexts where consumer law and the requirements of professional diligence on traders may be applicable. Furthermore, consumer law may be used to protect groups of consumers with vulnerabilities that have not been firmly established as protected characteristics.[58] This is an area where we have agreed to cooperate closely with the ICO and EHRC.

2.37 There are numerous examples in the literature of how algorithmic systems can give rise to potentially illegal discrimination,[59] including many that are

---

[54] Kleinberg, J, Ludwig, J, Mullainathan, S, & Sunstein, CR (2020), 'Algorithms as discrimination detectors', in *Proceedings of the National Academy of Sciences*, and *id*. (2018), 'Discrimination in the Age of Algorithms', *Journal of Legal Analysis,* Vol. 1, pp.113-174.

[55] The Centre for Data Ethics and Innovation's *Landscape Summary: Bias in Algorithmic Decision-Making* outlines several ways that industry and academia have come up with, such as statistical approaches and software toolkits, self-assessment tools, documentation standards, certification and auditing. It also emphasises that strategies for mitigating algorithmic bias depend on what a fair outcome looks like in a specific context. (Rovatsos, M et al. (2019), 'Landscape Summary: Bias in Algorithmic Decision-Making', Centre for Data Ethics and Innovation.)

[56] Data protection law could help mitigate risks of unfair and illegal discrimination. Data protection law, specifically the General Data Protection Regulation (GDPR), has played a significant role in spurring discussion. In addition to restricting the circumstances in which data controllers can carry out solely automated individual decision-making, the GDPR also requires data controllers to give individuals information about the processing, including an explanation of the decision and decision-making process. It also requires data controllers to take steps to prevent errors, bias and discrimination. (ICO, 'Rights related to automated decision making including profiling'.)

[57] Zuiderveen Borgesius, F (2018), 'Discrimination, artificial intelligence, and algorithmic decision-making', Council of Europe.

[58] For example, consumers that may have impulse-control disorders, such as Compulsive Buying Disorder, may be targeted by firms. (See, for instance, Gupta, S (2013), 'A literature review of compulsive buying–a marketing perspective', *Journal of Applied Business and Economics*, *14*(1), 43-48.) It is an open question on the extent to which businesses use algorithmic systems to identify and exploit consumers with impulse-control disorders.

[59] Many examples are provided in O'Neil, C (2016), *Weapons of Math Destruction*.

outside our potential remit such as criminal justice,[60] employee relations,[61] and health systems delivering healthcare services to populations.[62] There is also significant research into discriminative harms arising from algorithmic systems that provide information used to assist humans in decision-making, for example algorithmic systems that provide risk assessments of people looking to take out loans.[63] This "in-the-loop" decision making paradigm is not in scope for further consideration here.

2.38 The remainder of this section discusses selected examples to illustrate some economic and transactional contexts where indirect discriminatory outcomes can arise in algorithmic systems which affect: the outcomes that consumers experience, such as prices paid or quality of services received; the options that are available to them (discrimination on online 'sharing economy' platforms); and even the information that consumers have (discriminatory ad targeting).[64]

2.1.3.1 Geographic targeting

2.39 People with protected characteristics are unevenly distributed geographically. As a result, even simple policies implementing regional pricing or varying services available in different areas could potentially result in indirect discrimination.

2.40 For example, Larson et al. (2015) found that the price for an online SAT preparation service in the US varied substantially depending on where customers lived, and showed that Asians are almost twice as likely as non-Asians to live in areas offered higher prices.[65] In another example, when Amazon expanded free same-day delivery for its Prime service in 2016 in the US, it prioritised areas with high concentrations of Prime members. However, the effect of this was that predominantly black neighbourhoods were

---

[60] For instance, the debate around the COMPAS recidivism algorithm used in the US, which highlighted a fundamental tension between different conceptions of fairness in risk scoring. Kleinberg et al. (2016) showed that it is impossible to have risk scores which are well-calibrated within groups (e.g. white defendants and black defendants assigned the same risk score have the same probability of recidivism in aggregate) and balanced for the positive class (e.g. white and black recidivists have similar distribution of risks scores in aggregate), where there is imperfect prediction and base rates differ between groups. (Kleinberg, J, Mullainathan, S, & Raghavan, M (2016), 'Inherent trade-offs in the fair determination of risk scores', *arXiv preprint arXiv:1609.05807*.)
[61] Oppenheim, M (2018), 'Amazon Scraps 'Sexist AI' Recruitment Tool', *The Independent*, 11 October.
[62] Obermeyer, Z, Powers, B, Vogeli, C, & Mullainathan, S (2019), 'Dissecting racial bias in an algorithm used to manage the health of populations', *Science*, *366*(6464), 447-453.
[63] Green, B and Chen, Y (2019), 'The Principles and Limits of Algorithm-in-the-Loop Decision Making', in *Proc. ACM Hum.-Comput. Interact. 3*, CSCW, Article 50.
[64] Lee, T, Resnick, P, and Barton, G (2019) 'Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms', *Brookings*.
[65] Larson, J, Mattu, S, and Angwin, J (2015), 'Unintended Consequences of Geographic Targeting', *ProPublica*, 1 September, and Angwin, J, Mattu, S, and Larson, J (2015), 'The Tiger Mom Tax: Asians Are Nearly Twice as Likely to Get a Higher Price from Princeton Review', *ProPublica*, 1 September.

excluded, even (in some cases) where every area surrounding those neighbourhoods were served.[66]

2.1.3.2 Online sharing economy platforms

2.41 Platforms can make design choices which fail to mitigate discrimination arising from aggregate user behaviour, particularly where platforms attempt to build trust and facilitate transactions by reducing anonymity and providing information about people providing services.

2.42 TaskRabbit and Fiverr are online freelancing marketplaces that connect consumers to workers for small tasks. They show images of workers to customers who are deciding whether to engage them and who can leave reviews of the workers after they have performed the task. Hannák et al. (2017)[67] found that women and black workers received fewer reviews, with fewer positive adjectives and more negative adjectives in reviews, and lower feedback scores than other workers with similar attributes. In addition, the authors also analysed workers' rankings in the search algorithms and found significant gender and racial bias in rankings on TaskRabbit[68] (but not Fiverr), which may be partially explained if the algorithms take customer behaviour into account. Similarly, Edelman and Luca (2014) and Luca and Bazerman (2020) studied Airbnb, an online marketplace for short-term rentals that presents information about hosts and guests. The authors found that black hosts earn less, and black guests are rejected more.[69]

2.1.3.3 Discriminatory ad targeting

2.43 Online platforms can allow (or fail to prevent) advertisers targeting ads using protected characteristics. Angwin et al. (2017) found that it was possible to place housing ads which excluded anyone with an 'ethnic affinity' for African-American, Asian-American or Hispanic people on Facebook,[70] although Facebook has since made a number of modifications to its system to limit advertisers' ability to target by some demographic categories.

---

[66] Ingold, S and Soper, S (2016), 'Amazon Doesn't Consider the Race of Its Customers. Should It?', *Bloomberg*, 21 April.
[67] Hannák, A, Wagner, C, Garcia, D, Mislove, A, Strohmaier, M, & Wilson, C (2017), 'Bias in online freelance marketplaces: Evidence from Taskrabbit and Fiverr', in *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 1914-1933).
[68] However, the authors found that the specific groups that are ranked lower on TaskRabbit change from city-to-city.
[69] Edelman, BG, & Luca, M (2014) 'Digital discrimination: The case of Airbnb.com', *Harvard Business School NOM Unit Working Paper*, (14-054). See also Luca, M, and Bazerman, MH (2020), 'What data experiments tell us about racial discrimination on Airbnb', *Fast Company*.
[70] Angwin, J, Mattu, S, and Larson, J (2017), 'Facebook (still) letting housing advertisers exclude users by race', 21 November.

2.44 Even without explicit discriminatory intent by advertisers, it is possible that discriminatory outcomes can arise due to ad delivery systems optimising for cost-effectiveness (for example maximum impressions or conversions for lowest cost within budget). Ali et al. (2019) found that, despite setting equal and highly inclusive targeting parameters, there was significant skew in Facebook's ad delivery along gender and racial lines.[71] Ali et al. (2019) also found that campaigns with lower daily budgets showed fewer ads to women, confirming findings by Lambrecht and Tucker (2019) that, as there is more competition among advertisers to show ads to women, it is more expensive to do so. This can result in ad campaigns that were intended to be gender-neutral (e.g. for STEM career ads) being served in an apparently discriminatory way to more men because it is more cost-effective to do so when maximising the number of ad impressions for a given budget.[72]

2.1.4 Unfair ranking and design

2.45 Searching for a product or a service takes time and effort. Firms can provide value to consumers by aggregating, organising, and retrieving options that best meet consumers' needs. Well-designed choice architecture including default options and rankings can help consumers make decisions efficiently. If there is sufficient competition, informed and active consumers can switch to other platforms if they are unsatisfied with the results of one platform.

2.46 However, when firms are not transparent about the criteria they use to do so, this could give rise to suspicions or concerns that the default options and rankings may reflect what is in the firm's interest, potentially at the expense of consumers' interest.[73] This is especially the case where consumers are liable to misperceive defaults and ordered results as objective recommendations.

2.47 We define unfair ranking and design as the use of algorithmic systems to modify rankings or other design features to influence what a consumer sees to gain commercial advantage, but that ultimately degrades or misrepresents the offering to the consumer.

2.48 We discuss two important ways in which platforms may use unfair ranking and design:

---

[71] Ali, M, Sapiezynski, P, Bogen, M, Korolova, A, Mislove, A, & Rieke, A (2019), 'Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes', *arXiv preprint arXiv:1904.02095*.
[72] Lambrecht, A, & Tucker, C (2019), 'Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads', *Management Science*, *65*(7), 2966-2981.
[73] We acknowledge there are many good reasons for platforms not making algorithm criteria transparent, such as to prevent gaming or fraud.

(a) A platform may manipulate rankings of results to favour certain options, because it derives benefit from a commercial relationship, such as higher commission payments or revenue shares. (It may also favour options that it owns, which are competing against other options on the platform. This is known as self-preferencing, which we discuss further in the 'Self-preferencing' section below.)

(b) Platforms may use other unfair design practices ('dark patterns') to exploit consumers' behavioural biases for commercial gain, including the use of misleading scarcity messages, which exploit consumers' loss aversion and tendency to be influenced by the actions of others (social proof).

2.1.4.1 Preferencing others for commercial advantage

2.49 Platforms often have commercial relationships with other firms that participate on their platform. For example, platforms may charge a fee to firms to create listings, or the platform may receive commission or a share of the revenue from consumers that find or transact with firms listed on the platform. If these commercial relationships are differentiated, some firms may offer to pay more to the platform than other firms, in exchange for the platform giving more prominence to their options or otherwise distorting the ranking algorithm (relative to an algorithm that orders listings based 'competition on the merits' on factors that consumers value). If this is not transparent to consumers, this could be an unfair commercial practice. Moreover, a platform gain from unfair ranking is more likely outweigh any costs, from consumers perceiving the ranking to be lower quality and switching to an alternative, if it has greater market power and consumers less able to switch to an alternative.

2.50 The order in which results are presented matters because of ranking and ordering effects (or 'position bias'):[74] consumers are more likely to select options near the top of a list of results, simply by virtue of their position and independent of relevance, price or quality of the options.[75] These effects are even stronger on mobile devices, where an increasing proportion of online shopping is taking place,[76] and voice assistants, which are gaining in

[74] Position bias is closely related to consumer inertia and default effects, which are well-established behavioural biases where consumers tend to stick with defaults over demonstrably superior alternatives outside the default.
[75] See Finding 3 of the CMA (2017), 'Online search: Consumer and firm behaviour – A review of existing literature', 7 April. See also: Ursu, RM (2018), 'The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions', *Marketing Science,* 37 (4), 530 – 552; and De los Santos, B & Koulayev, S (2017), 'Optimising clock-through in online rankings with endogenous search refinement', *Marketing Science*, 36 (4), 542 – 564.
[76] Meola, A (2019), 'Rise of M-Commerce: Mobile Ecommerce Shopping Stats & Trends in 2020', *Business Insider*, 17 December.

popularity as the technology improves, as they present one or very few suggestions at a time.[77]

2.51 Firms can exploit default effects and ranking effects by placing options that are more profitable in prominent positions. This can be done in a way that is not transparent or understood and accepted by consumers, and potentially at the expense of the consumer if he or she would have chosen a superior alternative under a more neutral presentation of options. Where the favoured options belong to the same entity controlling the platform, this is a form of self-preferencing (which we discuss in the section on 'Self-preferencing' below). However, the favoured options need not belong to the platform. They could also be those of other firms with which the platform has some relationship (e.g. revenue sharing or commission).

2.52 A good example of this is the way some online hotel booking sites ranked search results for hotels, which was investigated by the CMA[78] and Australian Competition and Consumer Commission (ACCC). The CMA found that the search results on some hotel booking sites were affected by the amount of commission a hotel pays to the site. This was not made clear to consumers. The companies involved in the CMA investigation made commitments to be transparent about these practices.[79] Similarly, the ACCC found that Trivago breached Australian consumer law, as its ranking algorithm placed significant weight on which online hotel booking sites paid Trivago more, but misled consumers to believe that it provided an impartial, objective and transparent price comparison for hotel room rates.[80]

2.53 The Trivago example is analogous to undisclosed advertising or promoted content, where companies pay for placement. Even where advertising is disclosed, the disclosure can be imperceptible to consumers. On Google Search, for example, users can find it difficult to distinguish between advertisements and organic search results, suggesting that advertisements are insufficiently labelled.[81]

2.54 These unfair ranking practices are harmful to consumers, even without any personalisation of results to consumers. In a previous section, we explained how firms use various algorithmic systems, such as recommendation systems, ranking and information retrieval algorithms, to determine and

---

[77] Gearbrain (2020), 'Voice shopping with Amazon Alexa and Google Assistant: Everything you need to know', 11 March.
[78] CMA (2017), 'Online hotel booking: CMA launches consumer law investigation into hotel booking sites'. 27 October.
[79] CMA (2019), 'Hotel booking sites to make major changes after CMA probe.' 6 February.
[80] ACCC (2020), 'Trivago misled consumers about hotel room rates', 21 January.
[81] Lewandowski, D, Kerkmann, F, Rümmele, S & Sünkler, S (2017), 'An empirical investigation on search engine ad disclosure', *Journal of the Association for Information Science and Technology* 69(3).

optimise which alternatives to present, and the order in which to present them. We also explained how recommendation systems and ranking algorithms can generate personalised options and results pages for each consumer. Personalised rankings can further enhance the usefulness of results for consumers, but may also amplify any default and ranking effects that firms may be exploiting.

2.1.4.2 Dark patterns

2.55 Consumers have behavioural biases or vulnerabilities that can be exploited through different choice architecture. Dark patterns are user interface designs that trick users into making unintended and potentially harmful decisions.

2.56 Firms can deploy a vast array of different choice architectures and dark patterns, and algorithms are relevant for only a subset of these. Specifically, following the taxonomy of Mathur et al. (2019),[82] audits and analyses of firms' algorithmic systems would be useful for uncovering certain deceptive dark patterns.

2.57 We focus on the example of scarcity messages (promotional messages that highlight the limited availability of a product, either in quantity or in time). Scarcity messages can be generated by simple algorithmic systems that calculate the required metric. These messages can create a sense of urgency in consumers, and lead to buying more, spending less time to search, and improved opinion of the product.[83] However, scarcity messages must not be misleading or false.

2.58 To illustrate, the CMA investigated potentially misleading scarcity messages on hotel booking websites, such as "X other people are viewing this hotel right now" and "X rooms left". The CMA found that these claims could be incomplete. For example, the number of 'other people' viewing the hotel at times might include those looking at different dates or different room types, or there may be rooms at the hotel available on other platforms.[84]

2.59 More egregiously, we are aware of several blatant examples of ecommerce websites where these scarcity messages are completely fabricated. For

---

[82] Mathur et al. (2019) set out five characteristics that dark patterns may be categorised, in terms of how they affect user decision-marking: asymmetric, covert, deceptive, hides information, and restrictive. Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019), 'Dark patterns at scale: Findings from a crawl of 11K shopping websites', *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1-32.
[83] Cialdini, RB (1984), *Influence: The psychology of persuasion* (Vol. 55, p. 339), New York: Collins.
[84] See the discussion of the CMA's Hotel Online Booking case in Fung, S, Haydock, J, Moore, A, Rutt, J, Ryan, R, Walker, M, & Windle, I (2019), 'Recent Developments at the CMA: 2018–2019', *Review of Industrial Organization*, *55*(4), 579-605.

instance, users have reported, simply by inspecting the code for the relevant parts of some websites that feature scarcity messages, they can directly observe that the reported number of people viewing a product is a random number.[85]

2.60 We also note that A/B testing allows firms to further refine their choice architecture. Machine learning can be used to better achieve more granular segmentation that results in more personalised outcomes as discussed above. If machine learning or other automated systems orchestrate the A/B tests being conducted, the exact tests may lack appropriate oversight and could lead to inadvertent harm.

2.2 Exclusionary practices

2.61 In this section, we discuss algorithmic harms due to exclusionary practices: where algorithmic systems are used by dominant firms to deter competitors from challenging their market position.

2.2.1 Self-preferencing

2.62 For the purposes of this paper, we define self-preferencing as decisions by an online platform that favour its own products or services at the expense of those of its competitors.

2.63 Some of the most popular online marketplaces and search engines have strong market positions and represent a key gateway for businesses to access customers. The choice architecture and other design decisions made by those platforms, including the prominence of different options and products and how they are ranked, can have significant implications for the profitability of these businesses.

2.64 Competition can be harmed where a dominant platform favours its own products and services where they are in competition with rivals' products and services offered on its platform. This is particularly the case where the dominant platform's preference is not based on 'competition on the merits' between its own products and services and those of its rivals. In an online context, this may involve manipulating key algorithms and systems that operate their platforms, such as ranking algorithms, to favour their own products.

---

[85] Bergdahl, J (2020), 'Are 14 people really looking at that product?', 14 June. See also this Twitter thread by @Ophir Harpaz on 16 October 2019, which reports a similar example.

2.65 Google Shopping is a seminal case illustrating the role of algorithms in self-preferencing abuses. In that case, the European Commission found that Google had infringed competition law by positioning and displaying its own Google Shopping service in its general search result pages more favourably than competing comparison shopping services. Google had dedicated algorithms (Google's Panda update) that were designed to stop sites with poor quality content appearing in Google's top search results. A key finding in this case was that, whilst these algorithms reduced the ranking of competing comparison-shopping services and affected their visibility, Google exempted its own Google Shopping service from these algorithms and positioned it prominently in its general search pages.[86]

2.66 During our online platforms and digital advertising market study, we heard concerns from specialised search providers[87] about Google taking advantage of its market power in general search to self-preference its own specialised search products or foreclose specialised search competitors. These concerns included Google taking various design decisions that direct traffic to its own specialised search products; demoting links to its specialised search rivals' organic search results pages; and extracting rent from specialised search providers by inflating their cost of search advertising and increasing the prominence of ads at the expense of organic links.[88]

2.67 To take another example, in 2019, the Wall Street Journal alleged that Amazon had changed its search algorithm to more prominently feature listings that are more profitable for Amazon. Instead of showing customers mainly the most relevant and best-selling listings when they search, the change allegedly benefited Amazon's own private label products on its platform at the expense of competing products on Amazon Marketplace.[89]

2.68 Self-preferencing is not limited to ranking. Another example of alleged self-preferencing that potentially exploits default effects or saliency is the way in which Amazon selects which seller occupies its Featured Offer (which we also refer to by its former and more well-known name, the 'Buy Box'), the default retailer for any product listed on Amazon. The Featured Offer is the offer near the top of a product detail page, which customers can buy now or add to their

---

[86] European Commission decision of 27 June 2017, Case AT.39740 – Google Search (Shopping).
[87] Specialised search, sometimes described as vertical search, provide tools that allow consumers to search for, compare, and purchase products or services from different providers in a particular sector (or 'vertical') such as travel, local search, consumer finance, etc.
[88] These concerns and evidence are presented in Appendix P of CMA (2020), 'Online Platform and Digital Advertising Market Study final report'.
[89] Wall Street Journal (2019), 'Amazon Changed Search Algorithm in Ways That Boosted Its Own Products', 16 September. It further alleged that the algorithm uses 'proxies' for profitability, variables that correlated with improved profitability for Amazon, but which an outside observer may not be able to tell that, after Amazon lawyers rejected an early proposal to 'add profit directly into the algorithm' because of concerns that it would 'create trouble with antitrust regulators'.

shopping carts. A high proportion of sales on Amazon occur through the Buy Box, with estimates ranging from 80 percent to 90 percent.[90] The precise details of the algorithm that Amazon uses to select sellers to fulfil orders made by the Buy Box (which could be Amazon itself) are not public. However, Amazon explains that the chance of winning the Buy Box is affected by the price, availability, fulfilment method and customer experience metrics. Third-party sellers on Amazon Marketplace worry that Amazon's Buy Box algorithm unfairly favours Amazon's own products. In July 2019, the European Commission opened a formal investigation into Amazon and will look at the role of data in selecting the winners of the Buy Box and the impact of Amazon's potential use of competitively sensitive marketplace seller information on that selection.[91]

2.69 Going beyond the potential misuse of choice architecture to influence the decisions of consumers and customers, firms using algorithmic systems may do so in situations where there are information asymmetries or potential conflicts of interest. For instance, in some cases, consumers and customers effectively delegate decisions to firms who act on their behalf using an algorithmic system. In other cases, a platform can operate an exchange, and act as an intermediary for two or more customers. Where these firms' and platforms' behaviour are implemented through automated decision systems or algorithms, these algorithms can be central to the functioning of those markets. If there is insufficient trust and transparency, there may be concerns that the firm or platform is making decisions that benefit itself rather than those it is supposed to act for, and this could result in harms to consumers and competition.

2.70 For example, in our Online Platforms and Digital Advertising market study, we heard concerns that Google is able to use its control of key parts of the adtech stack to determine auction processes in a way which favours its own digital advertising businesses.[92] We also heard that there was a lack of transparency of algorithms used in advertising auctions, including those used to weight bids by relevance and automated bidding algorithms.[93]

2.71 Information asymmetries and lack of trust and confidence in the integrity of the operations of key algorithms can lead consumers and customers to stop participating in digital markets, for example quitting social media apps or

---

[90] For instance, RepricerExpress states that '83% of all Amazon sales happen through the Buy Box and even more on mobile'. (RepricerExpress (2020), 'How to Win the Amazon Buy Box in 2020').
[91] European Commission (2019), 'Antitrust: Commission opens investigation into possible anti-competitive conduct of Amazon', [Press release] 17 July; European Commission (2020), 'Commission opens second investigation into its e-commerce business practices', [Press release] 10 November 2020.
[92] See Chapter 5 and Appendix M of CMA (2020), 'Online Platform and Digital Advertising Market Study final report'.
[93] Chapter 5, Appendices Q and U, *ibid*.

stopping using Google. This is particularly the case in the absence of strong enough incentives on firms – such as bolstering reputation – to provide more information to consumers. Where firms and market participants are unable to overcome these problems, it may be useful for regulators to intervene and help alleviate information asymmetries by directly investing in capabilities and efforts to audit and check how these algorithms operate, or to help facilitate the development of independent algorithmic audit providers. These ideas are discussed in more detail in the sections below.

2.2.2 Manipulating platform algorithms and unintended exclusion

2.72 Manipulation of ranking algorithms and other platform design choices could also allow incumbents on a platform to exclude competitors. Third-party incumbents selling on online marketplaces may be able to game the algorithms of key platforms, in order to temporarily suppress the visibility and prominence of new entrants at a critical juncture for the entrant's growth. More generally, the search and ranking algorithms of key platforms may place insufficient weight on maintaining competition and reinforce market power in other markets. For instance, platforms may systematically favour established incumbents by making it too difficult for consumers to discover and potentially switch to a new entrant, and may even extract a share of the rents accruing to incumbent firms due to lack of competition in other markets.

2.73 More generally, changes to complex algorithmic systems for legitimate aims can create unintended harms to businesses that rely on them and harm competition in affected markets. In particular, changes to the algorithms of gateway platforms[94] such as Facebook and Google can have substantial effects, and many businesses may be inadvertently excluded or marginalised as a result. Understanding and adapting to changes in the operation of these algorithms can be a barrier to entry.

2.74 For example, during the CMA's market study on Online Platforms and Digital Advertising, many publishers expressed concern about unexpected and unexplained changes to search and ranking algorithms, most notably in relation to Google Search[95] and Facebook News Feed.[96] Publishers argued

---

[94] The Furman Review ('Unlocking digital competition, Report of the Digital Competition Expert Panel', March 2019) introduces the term 'competitive gateway' to describe platforms that have a position of control over other parties' market access.

[95] In 2019, Google changed its news search algorithm to highlight 'significant original reporting', although there is no objective definition for 'original reporting'. (Gingras, R (2019), 'Elevating original reporting in Search'. Google. 12 September.) Google's changes could potentially change the business models of hundreds of websites. (Bohn, D and Hollister, S (2019), 'Google is changing its search algorithm to prioritize original news reporting'. *The Verge*. 12 September.)

[96] For example, a change to Facebook's curation algorithm in 2016 to show users more stories from friends and family resulted in a significant decline in traffic to publishers' sites, up to 25 percent for some. See Isaac, M and Ember, S (2016), 'Facebook to Change News Feed to Focus on Friends and Family', *The New York Times*, 29

that a reduction in website traffic resulting from an algorithm change has direct financial consequences for their businesses, and that sudden, unexplained and significant algorithm changes make planning and financial decision-making more complicated. Businesses incur significant costs understanding and optimising content to adapt to these changes. Some publishers have also told us that they think that in some cases algorithm changes may be commercially motivated to favour the platforms or affiliated parties at the expense of other publishers.[97]

2.75 Potentially affected businesses should receive sufficient explanation of how these algorithms work and sufficient notice of changes that may impact them significantly. We discuss the importance of explainability and transparency in Sections 3 and 4. Also, as emphasised in our discussion where harms may be unintended (such as section 2.3 on algorithmic collusion, and section 2.1.3 on algorithmic discrimination), firms are responsible for the effects of their algorithmic systems, including unintended effects.

2.2.3 Predatory pricing

2.76 We discussed some potential harms to consumers from personalised pricing algorithms in section 2.1.2 above. This included how firms may can use machine learning and data science techniques to analyse which customers are likely to switch.

2.77 In theory, it is possible that incumbent firms may use similar data, algorithms and techniques for personalised pricing in order to identify and selectively target those customers most at risk of switching, or who are otherwise crucial to a new competitor. This could make it easier, more effective, and less costly for incumbent firms to predate successfully,[98] and to foreclose or marginalise competitors.

2.3 Algorithmic collusion

2.78 In this section, we discuss the use of algorithmic systems to facilitate collusion and sustain higher prices. Algorithms could be used to make explicit collusion

---

June, and Tynan, D (2016), 'Facebook's newest news feed: good for friends, bad for publishers', *The Guardian*, 29 June.

[97] CMA (2020), 'Online Platforms and Digital Advertising market study final report, Appendix S', paragraphs 21 to 31.

[98] Predation refers to situations where an incumbent firm with a dominant position sets prices very aggressively with the aim of excluding a rival from the market. If successful, the predator will be able to recoup its losses by raising prices and earning higher profits because the prey is longer in the market. Predation is controversial because it is difficult to distinguish low prices due to tough but fair competition from low prices that are part of an exclusionary strategy by a dominant incumbent.

more stable or could collude without any explicit agreement or communication between firms.

2.79 The potential use of pricing algorithms to facilitate collusion has received much attention from both academics and competition regulators. In 2016, Ezrachi and Stucke published a book setting out potential collusive scenarios[99] and, as mentioned above, in 2018, we published an economic working paper on this topic. Since then, several other competition authorities have also published reports,[100] and academic researchers have published notable papers that show that pricing algorithms can, in theory, learn to collude.

2.80 Broadly, the concerns around algorithmic collusion fall into three categories:

(a) First, the increased availability of pricing data and the use of automated pricing systems can **facilitate explicit coordination**, by making it easier to detect and respond to deviations and reducing the chance of errors or accidental deviations. Even simple pricing algorithms, with access to real-time data on competitors' prices, could make explicit collusion between firms more stable.

(b) Second, where firms use the same algorithmic system to set prices, including by using the same software or services supplied by a third-party, or by delegating their pricing decisions to a common intermediary, this can create a '**hub-and-spoke**' structure and facilitate information exchange.

(c) Finally, there is a possibility of '**autonomous tacit collusion**', whereby pricing algorithms learn to collude without requiring other information sharing or existing coordination.

2.3.1 Facilitate explicit coordination

2.81 There have been a few enforcement cases by competition authorities against firms that used pricing algorithms to enforce explicit collusive agreements, such as the online posters case (the CMA's Trod/GB eye decision and US v. Topkins).[101] We are also aware of an investigation by the CNMC into Spanish real estate intermediation franchises and suppliers of IT solutions for real

---

[99] Ezrachi, A and Stucke, ME (2016), 'Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy'. *Harvard University Press*.
[100] These include: Autorité de la Concurrence and Bundeskartellamt (2019), 'Algorithms and Competition'; and Autoridade da Concorrência (2019), 'Digital ecosystems, Big Data and Algorithms – Issues Paper'.
[101] US Department of Justice (2015), 'Former E-Commerce Executive Charged with Price Fixing in the Antitrust Division's First Online Marketplace Prosecution' [Press release] 6 April. CMA (2016), 'Online seller admits breaking competition law' [Press release] 21 July.

estate brokerage. The allegations are that these undertakings have facilitated coordination of prices and other terms of sale by real estate agents, through the design of brokerage software and algorithms.[102,103]

2.82 Algorithms may also be used to enforce anti-competitive restraints more generally, to set and maintain supra-competitive prices (i.e. prices that are higher than would be sustained in a competitive market). For example, in 2019, an online travel agency alleged that its competitors Booking.com and Expedia were de facto enforcing wide price parity clauses by demoting in their rankings any hotel operator that offers cheaper rooms on rival booking websites, so that they are less likely to appear or appear lower in their search results.[104] These allegations are consistent with recent academic work suggesting that online travel agents alter their search results to discipline hotels for aggressive prices on competing channels, and that this reduces search quality for consumers.[105] Online travel agents' wide price parity clauses have been the subject of numerous investigations by EU competition authorities,[106] and Booking.com and Expedia made formal commitments, including to the CMA,[107] not to enforce wide price parity clauses.[108]

2.3.2 Hub-and-spoke

2.83 Another potential concern is the extent to which pricing recommendations or price setting by common intermediaries could result in supra-competitive prices.[109] Many platforms offer tools and algorithms to their supply-side users (such third-party sellers on Amazon Marketplace and eBay and hosts on Airbnb), in order to help them to set and manage their prices. (For example, Amazon provides Automate Pricing for its third-party sellers.) Some sharing economy platforms go further and recommend prices, allow supply-side users to delegate pricing to the platform, or even to require them to do so. It is an

---

[102] CNMC (2020), 'The CNMC opens antitrust proceedings against seven firms for suspected price coordination in the real estate intermediation market', [Press release] 19 February.
[103] Arguably, the Eturas case concerning a decision by the Lithuanian Competition Council against a common online travel booking system (Eturas) which facilitated collusion by restricting discounts can also be viewed as an 'algorithmic' collusion case, in the limited sense that the discount cap was implemented using technical restrictions on travel agencies using the Eturas system. (Case C-74/14, Eturas UAB and Others v Lietuvos Respublikos konkirencijos taryva (Eturas), ECLI:EU:C:2016:42.)
[104] Global Competition Review (2019), 'Expedia and Booking.com accused of imposing "new type" of abusive price parity', 13 June.
[105] Hunold, M, Reinhold, K, & Laitenberger, U (2018), 'Hotel Rankings of Online Travel Agents, Channel Pricing and Consumer Protection', Discussion Paper, *Duesseldorf Institute for Competition Economics*.
[106] European Competition Network (2016), 'Report on the monitoring exercise carried out in the online hotel booking sector by EU competition authorities in 2016'.
[107] CMA (2020), 'Voluntary extension to parity commitments by Booking.com and Expedia', 20 August.
[108] It is well-established that wide parity clauses (also known an 'most favoured nation' clauses) breach competition law. See CMA (2020), 'CMA fines ComparetheMarket £17.9m for competition law breach', press release.
[109] See paragraphs 5.19 to 5.21 of CMA (2020), 'Pricing algorithms'.

open question whether these platforms' algorithms optimise prices and recommendations for each user independently.

2.3.3 Autonomous tacit collusion

2.84 Some of the concerns about algorithmic collusion have focused on the prospect of complex and sophisticated pricing algorithms – using techniques like deep reinforcement learning – learning by themselves to tacitly collude without explicit communication and intention by human operators to suppress rivalry.[110]

2.85 Simulation studies show that there are clear theoretical concerns that algorithms could autonomously collude without any explicit communication between firms. For example, Calvano et al (2019) showed that Q-learning (a relatively simple form of reinforcement learning) pricing algorithms competing in simulations can learn collusive strategies with punishment for deviation, albeit after a number of iterations of experimentation in a stable market.[111]

2.3.4 Empirical evidence

2.86 The extent to which more sophisticated pricing algorithms are used across real-world markets is uncertain, with a wide range of estimates. Chen at al. (2016) analysed third-party sellers on Amazon Marketplace and found 543 sellers (out of around 33,300 unique seller IDs in their crawled datasets) that they regarded as very likely to be using pricing algorithms.[112,113] In May 2017, the European Commission's E-commerce Sector Inquiry found that approximately 28 percent of respondents use software to track and subsequently adjust their own prices.[114] In 2019, the Portuguese competition regulator (AdC) found approximately 8 percent of firms they surveyed used pricing algorithms.[115] These numbers are likely to be increasing as third-party pricing software is provided by an increasing number of firms and becomes

---

[110] See, for instance, the discussion in Chapter 8 of Ezrachi, A and Stucke, ME (2016), 'Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy'. *Harvard University Press*.
[111] Calvano, E, Calzolari, G, Denicolo, V, and Pastorello, S (2019), 'Artificial Intelligence, Algorithmic Pricing and Collusion'; and Klein, T (2019), 'Autonomous Algorithmic Collusion: Q-Learning Under Sequential Pricing'.
[112] Chen, L, Mislove, A, & Wilson, C (2016), 'An empirical analysis of algorithmic pricing on Amazon Marketplace' in *Proceedings of the 25th International Conference on World Wide Web* (pp. 1339-1349). Surprisingly, as the authors themselves noted, third-party algorithmic sellers in that study sell fewer unique products by a large margin, suggesting that they tend to specialise in a relatively small number of products.
[113] In previous work (CMA 2018), we spoke with a number of providers of algorithmic pricing tools and services, and they indicated that larger third-party sellers on Amazon (e.g. those exceeding $1m annual revenue) tend to have automated repricing software in order to manage large numbers of products (CMA (2018), 'Pricing Algorithms: Economic Working Paper on the Use of Algorithms to Facilitate Collusion and Personalised Pricing').
[114] The European Commission found that 53% of the respondent retailers track the online prices of competitors, and 67% of them also use automatic software for that purpose. 78% of those retailers that use software to track prices subsequently adjust their own prices. (European Commission, 'Final Report on the E-commerce Sector Inquiry', Commission Staff Working Document, May 2017.)
[115] Autoridade da Concorrência (2019) 'Digital Ecosystems, Big Data and Algorithms: Issues Paper'.

available to more businesses who have previously lacked the internal capabilities to do this.

2.87 In general, the risks of collusion in real-world markets is unclear due to a relative paucity of empirical evidence. As discussed above, there have been few enforcement cases by competition authorities against firms that used pricing algorithms to enforce explicit collusive agreements. It is as yet unclear that competition authorities can object to hub and spoke and autonomous tacit collusion situations where, for example, there may not have been direct contact between two undertakings or a meeting of minds between them to restrict competition.

2.88 One recent academic study, Assad et al. (2020) is (to our knowledge) the first empirical analysis of the relationship between algorithmic pricing and competition in a real-world market. They estimated that German retail petrol stations increased their margins by around 9 percent after adopting algorithmic pricing, but only where they faced local competition. The margins do not start to increase until approximately a year after market-wide adoption, suggesting that algorithms in this market have learnt over time to coordinate on a tacit collusion outcome.[116]

2.4 Ineffective platform oversight harms

2.89 We define ineffective platform oversight harms as the stated use of algorithms to address harms that may be partially or wholly ineffective in practice that, if accompanied by a lack of transparency, cannot be externally evaluated. This can result in consumer harms (for example, businesses not being incentivised to improve their services), as well as wider social harms. This harm could cut across many of the other harms, but we highlight it separately due the importance for regulators in addressing lack of transparency.

2.90 An example of this is the harm caused by ineffective algorithms that are designed to combat fake online reviews. Consumer group Which? has found thousands of fake or suspicious reviews for popular technology products from 'unknown' brands for sale on Amazon.[117] Which? reported that, according to Amazon's own estimates, the majority of these are generated by computers. Its machine learning algorithms analyse all incoming and existing reviews, and block or remove those that are identified as inauthentic. However, Which? has found that many still get through these filters. This harm may be mitigated

---

[116] Assad, S, Clark, R, Ershov, D, & Xu, L (2020), 'Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market, *CESifo Working Paper*, No. 8521.
[117] Walsh, H (2019), 'Thousands of 'fake' customer reviews found on popular tech categories on Amazon', *Which?,* 16 April.

by the fact that platforms should be incentivised to make their algorithms more effective, since, if fake reviews are widespread, the platforms may lose consumer confidence and lose business to other platforms. However, where products must be purchased before a review can be posted, platforms may benefit from fake reviewing. The CMA's own work on fake online reviews identified, for example, that more than three-quarters of people are influenced by reviews when they shop online. The failure to detect these reviews and remove them can lead to consumers purchasing products or services that they do not want.

2.91 Another example is the harm caused by algorithms that are ineffective in filtering out harmful content.[118] In November 2019, Facebook announced that it was able to remove seven million instances of hate speech in the third quarter of 2019, 59 percent higher than the previous quarter.[119] Most of this is being detected by algorithms. It is unlikely that an algorithm would be able to capture all instances of hate speech in any given context, given the evolving dynamics of social topics that algorithmic systems are required to react to. Nonetheless, the hate speech classifier algorithms that Facebook used only work in around 40 languages, with Facebook citing insufficient training data on which to train algorithms for lesser-spoken languages. For many other algorithms used by large technology companies to filter out undesirable content, it remains unclear how effective they are. This results from a lack of transparency, which these companies are not compelled to provide to regulators or consumers.

2.92 There is an ongoing debate around platforms' liability for user-generated content uploaded on their platform. In the United States, Section 230 of the 1996 Communications Decency Act broadly provides immunity for platforms to content posted by users. It does, however, still hold companies liable for content that violates intellectual property law or criminal law.[120] From a UK consumer protection perspective, we expect platforms to ensure that their systems can detect illegal or misleading content. Further, Ofcom is set to become the regulator for online harms. In the Government's initial response to the consultation on the Online Harms White Paper in February 2020, it emphasised that companies would be required to fulfil their duty of care. This duty of care includes removing illegal content, and for content that is legal but has the potential to cause harm, stating publicly 'what content and behaviour

[118] Online harms such as hate speech are significant topics in themselves, however they are not covered extensively in this paper as they are being addressed through online harms regulation and Ofcom.
[119] Perrigo, B (2019)c, 'Facebook Says It's Removing More Hate Speech Than Ever Before. But There's a Catch', *TIME*, 27 November.
[120] Reuters (2019), 'Google and Reddit defend law protecting tech companies from liability for user-generated content', *Venture Beat,* 16 October.

they deem to be acceptable on their sites and enforce this consistently and transparently.'

3. Techniques to investigate these harms

3.1 In this section, we discuss techniques to investigate how widespread or problematic the harms outlined in the previous section are. We consider separately techniques that can be used without any access to companies' data and algorithms and those that require information from companies available only to employees, and possibly auditors including regulators. Some of the techniques will be more applicable to investigating certain harms than others.

3.1 Techniques to investigate harms without direct access to firms' data and algorithms

3.2 Without direct access to a firm's data or algorithms, it can be difficult to analyse a specific algorithm in isolation from the broader product or automated system in which the algorithm is deployed. Instead, most likely one would analyse the inputs and outputs of a system understood to be powered by one or more algorithmic processes.[121] A large set of the harms outlined earlier in this paper can be analysed by collecting or simulating appropriate data for use as input to a given algorithmic system, and then analysing the output; for example through conducting a systematic analysis of inputs and outputs through what the Ada Lovelace Institute and DataKind UK call a 'bias audit'.[122]

3.3 However, harms from an algorithmic service are best understood with the full and proper context of how such a service or automated system was designed, developed and trained, the data used as inputs, and how consumers respond to and make use of the output of the system.[123] Where a given algorithm does not have clear and transparent inputs and outputs, it can be difficult to understand the behaviour of the algorithm, rather than the wider automated

---

[121] The algorithm output, of say a search algorithm, might subject to additional effects such as noise, A/B testing or personalisation that affect the exact "search results" that a consumer sees. There is no guarantee that if two separate parties search for the same term at the same time, that they will see the same results in the same order. E.g. Hannak, A, Soeller, G, Lazer, D, Mislove, A and Wilson, C (2014), 'Measuring Price Discrimination and Steering on E-commerce Web Sites' in *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*, Association for Computing Machinery, New York, NY, USA, 305–318.

[122] This type of audit tests a hypothesis by looking at inputs and outputs to detect any biases in the outcome of a decision, as opposed to a broader inspection to investigate whether a system is compliant with specific regulations. See Ada Lovelace Institute & DataKind UK (2020), 'Examining the Black Box: Tools for Assessing Algorithmic Systems'.

[123] An example of a framework for understanding this broader context is in the Turing Institute's 2019 report on "Understanding Artificial Intelligence Ethics and Safety" – they outline four complementary categories of "Fairness": Data, Design, Outcome and Implementation.

system. Earlier in this paper we discussed how algorithms may enrich user-provided input with additional information, such as the behavioural history of a user, their device type or geo-location. This can lead to a set of algorithm inputs that are opaque and partially or entirely unobservable. For example, recommender systems typically train on a large collection of consumers' behavioural histories,[124] but exactly which behaviours are used as inputs can be unclear to an external party. Algorithmic outputs can also be unobservable, for example clustering methods that perform customer segmentation. Individual customers will often be unaware both that such segmentation is happening and which segment they may have been assigned to.[125]

3.4 Certain classes of harm described in this paper are generated as a result of interactions between multiple automated systems or algorithms, for example algorithmic collusion. In these instances, it might be possible to identify "aggregate" harm that results from the full set of algorithmic interactions, for example a significant reduction in price competition regarding a given product. It will, however, likely be much more challenging to understand the role of each individual algorithm in manifesting such a harm.[126]

3.5 While we recognise these potential shortfalls, we still believe that there is significant value in investigating automated systems without direct access to the underlying code, and note that academics have developed methods that are successful at disentangling algorithmic output from the wider system's output. For example, for price discrimination on e-commerce website search results, Hannák et al. (2014) developed a methodology to isolate the effects of personalisation from noise, and to measure the relevance of factors such as choice of operating system or purchase history.[127]

3.6 Where algorithmic systems have transparent outputs, one traditional methodological approach is to enlist consumers to act as digital "mystery shoppers". The CMA has used this technique in the past to understand how consumers use digital comparison tools when making a purchase through a

---

[124] Bobadilla, J, Ortega, F, Hernando, A, Gutiérrez, A (2013), 'Recommender systems survey', *Knowledge-Based Systems*, Volume 46, 2013, Pages 109-132, ISSN 0950-7051,
[125] Yan, J, Liu, N, Wang, G, Zhan, W, Jiang, Y, and Chen, Z (2009), 'How much can behavioral targeting help online advertising?' in *Proceedings of the 18th international conference on World wide web*, Association for Computing Machinery, New York, NY, USA, 261–270.
[126] For example, consider a set of pricing algorithm "agents" competing in the same market and assume these algorithms use other agents' prices and consumer demand to inform their own price. It is extremely challenging to simulate market conditions across possible market conditions to identify when harms may occur and identify which set of algorithmic interactions are responsible. This framework is the same as that in which the 2010 US Stock Market "Flash Crash" occurred; an event generally accepted as hard to have predicted *ex ante* and notoriously challenging to understand *ex post,* with many expert parties still disagreeing over the root cause – see this SEC report on the Flash Crash, and this more general 2018 FCA Report on Algorithmic Trading Compliance.
[127] Hannák, A, Soeller, G, Lazer, D, Mislove, A, and Wilson, C (2014), 'Measuring Price Discrimination and Steering on E-commerce Web Sites', in *Proceedings of the 2014 Conference on Internet Measurement Conference*, Association for Computing Machinery, New York, NY, USA, 305–318.

website or app. The European Commission has also used this approach in a consumer market study on online market segmentation through personalised pricing,[128] and the CMA has undertaken a similar study.[129] Such studies collect and use information such as real shoppers' online profiles, click and purchase history, device and software usage to identify possible harms.[130] Researchers have also performed studies on search result rankings, observing how they differ across devices (which can be correlated with protected characteristics), or when searches are repeated.[131] They collected historical data by running a variety of search queries and recording and comparing results[132] to identify possible harms, such as on different online price comparison sites.[133]

3.7 Outputs can also be observed through what Sandvig et al. (2014) call a 'scraping audit'.[134] Crawling and scraping are methods that allow data to be extracted from websites.[135] For example, Hannák et al. (2017) use scraping to extract data (such as demographic data, ratings and reviews, and workers' rank in search results) from two freelance marketplaces to create worker profiles and investigate the presence of bias in perceived gender or race.[136] Scraping to retrieve the output of a system can also be used in conjunction with real data inputs from the team undertaking the audit, for example to increase transparency around what data has been used to produce a personalised recommendation or price.[137] Lécuyer et al. (2014) propose a personal data tracking system for the Web that predicts, through correlation, which input data, such as a personal email or web search, are being used for targeting outputs, such as prices on an online marketplace.[138] More generally,

---

[128] European Commission (2018), 'Consumer market study on online market segmentation through personalised pricing/offers in the European Union'. 19 July. ISBN 978-92-9200-929-8
[129] Annex 1 of CMA (2018), "Pricing algorithms - Economic working paper on the use of algorithms to facilitate collusion and personalised pricing'.
[130] *ibid*.
[131] It should be noted that testing inputs and outputs should match how the algorithm works in practice, where feasible, as it may be challenging to rule out variations in outputs based on personalisation or user history.
[132] Fletcher, A (2019), 'Ranking rankings: Or how to protect consumers from being misled when searching online', *CCP Annual Conference*.
[133] Hunold, M, Reinhold, K, & Laitenberger, U (2018), 'Hotel Rankings of Online Travel Agents, Channel Pricing and Consumer Protection', Discussion Paper, *Duesseldorf Institute for Competition Economics*.
[134] Sandvig, C, Hamilton, K, Karahalios, K & Langbort, C (2014), 'Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms', *Paper presented to "Data and Discrimination," a pre-conference of the 64th annual meeting of the International Communication Association, Seattle, WA, USA*.
[135] Some websites have Terms of Service that prohibit crawling and scraping; however, we support their use when deployed in a responsible way for auditing purposes.
[136] Hannák, A, Wagner, C, Garcia, D, Mislove, A, Strohmaier, M, & Wilson, C (2017), 'Bias in Online Freelance Marketplaces: Evidence from TaskRabbit and Fiverr' in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1914-1933).
[137] Lécuyer, M, Ducoffe, G, Lan, F, Papancea, A, Petsios, T, Spahn, R, Chaintreau, A, & Geambasu, R (2014), 'XRay: Enhancing the Web's Transparency with Differential Correlation', in *Proceedings of the 23rd USENIX Conference on Security Symposium*.
[138] The authors create a plugin that collects the chosen inputs and outputs to be tracked, as well as 'shadow accounts' that use a subset of the input data from each personal account. A correlation engine then determines the strength of the association between the inputs and outputs. If an output appears highly correlated with an

we caution that there are no guarantees that a 'scraping audit' is either feasible or straightforward for a specific case, and this can depend on the nature of how a particular website or service operates.

3.8 More complex, technical approaches can include the use of APIs,[139] reverse engineering services to allow direct testing, or emulating web applications to submit queries with specific data and to intercept the complete output. The last two data collection methods were used by Chen et al. (2015) to audit Uber's surge price algorithm.[140]

3.9 The process by which relevant input data is collected and prepared for the investigation of an algorithm requires consideration from a practical and ethical viewpoint. It is important to consider the legal and ethical implications of data collection techniques such as web scraping or reverse engineering, as well as ensuring that we don't subject audited companies to undue burdens. Regarding handling any personal data, such collection, processing and storage must be performed in an ethical manner and in line with data protection law. In preparing such data, it is also important to be aware of introducing bias which may distort later analysis of the algorithmic service. For example, when testing image recognition algorithms for racial bias, one approach to data preparation is to use "off the shelf" algorithms to label race categories on a collected set of facial images. Those using these algorithms must be wary of whether the algorithms that perform such labelling suffer from similar biases to the ones we might be trying to audit.[141]

3.10 In the absence of accessible real-life data, some institutions are creating fake personae that simulate users of various demographics and interests. Such personae allow the use of automated methods to analyse certain discriminative harms at a larger scale than can be done manually. The Princeton WebTAP programme has developed OpenWPM, an open-source software that enables researchers to build up a history of browsing activity and observe how these personae interact with a digital service.[142] Datta et al. (2015) have also created a tool, AdFisher, that creates simulated user web histories, segmented by gender and age, and compares advertisements

---

input in several shadow accounts, it is possible to identify the input that is more likely to be responsible for that output, such as clicking on a product and later seeing an advert for it.

[139] An application programming interface (API) is a computing interface that allows data to easily be sent back and forth between systems (including inputs and outputs to an algorithmic system).

[140] Chen, L, Mislove, A, & Wilson, C (2015), 'Peeking beneath the hood of Uber', in *Proceedings of the 2015 internet measurement conference* (pp. 495-508).

[141] For the inference of gender and ethnicity from profile pictures, mainstream datasets and algorithms trained using these are known to suffer from biases, which is the reason Karkkainen and Joo (2019) put together the FairFace dataset. Kärkkäinen, K, & Joo, J (2019), 'Fairface: Face attribute dataset for balanced race, gender, and age', *arXiv preprint* arXiv:1908.04913.

[142] Narayanan, A & Resiman, D (2017), 'The Princeton Web Transparency and Accountability Project', *Transparent data mining for Big and Small Data*.

shown to each segment to analyse any indication of discrimination, or lack of transparency or choice in advertisement settings.[143] However, it is important we realise that "simulated" personae are not the same as real consumers, whose broader characteristics and behavioural histories may vary in ways that have material effects on the algorithms we wish to analyse.[144]

3.11 Once output has been collected, it requires appropriate analysis to test whether any hypothesised harm is indeed present. The most appropriate analytical methodologies depend on the theory of harm being tested, the system being investigated and the domain and context of how the system is deployed. Even when investigations can be framed in terms of statistical analyses, moral and legal aspects may still need to be considered. The growing corpus of ethical Machine Learning literature notes that there are varied definitions and interpretations of what constitutes 'fair', which themselves are complex and context specific,[145] and furthermore, the techniques to detect and remedy fairness concerns often require access to protected characteristics data that is often not readily available.[146]

3.2 Techniques to investigate harms when direct access to the data and algorithm is possible

3.12 An algorithm is developed by writing code which often, but not always, undergoes optimisation routines that process relevant data to produce a final, "trained" algorithm. Having access to the data and/or the code means it is possible to audit a decision-making system through a comprehensive regulatory inspection in a more thorough manner than a "black-box" approach.[147] It may not be necessary to collect and process the data; it is also possible to run the algorithm code on test data and directly analyse its outputs and limitations. Likewise, we may not necessarily need access to code; relevant data from the algorithm owner, such as inputs and outputs from live deployment or A/B testing, in itself might be sufficient for certain audit purposes.

---

[143] Datta, A, Tschantz, M, & Datta, A (2015), 'Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice and Discrimination', in *Proceedings on privacy enhancing technologies.*

[144] An example is Deep Reinforcement Learning Algorithms, which are known to be quite sensitive to minor changes in inputs and training – see: Henderson, P, Islam, R, Bachman, P, Pineau, J, Precup, D, Meger, D (2017), 'Deep Reinforcement Learning that Matters'. *Arxiv preprint.*

[145] See, for example, Binns, R (2018), 'Fairness in machine learning: Lessons from political philosophy' in *Conference on Fairness, Accountability and Transparency* (pp. 149-159). PMLR, or Hutchinson, B, & Mitchell, M (2019), '50 years of test (un) fairness: Lessons for machine learning', in *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 49-58).

[146] Veale and Binns discuss three broad options for mitigating discrimination without access to protected characteristics data in Veale, M, & Binns, R (2017), 'Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data', *Big Data & Society*, *4*(2).

[147] Ada Lovelace Institute & DataKind UK (2020), 'Examining the Black Box: Tools for Assessing Algorithmic Systems'.

3.13 Where there is access to the code, there are three possibilities for investigation: "dynamic analysis", e.g. automated testing through execution of the code; "static analysis" e.g. identifying format errors, where the code can only be analysed in isolation from its environment; and a manual "code review", which would be extremely challenging for complex algorithmic software. The latter two approaches have noted shortcomings, such as the inability to understand how the code's external dependencies affect its behaviour. Additionally, the continued existence of malware in fully transparent open-source software illustrates how harmful behaviours can perpetuate unseen.[148] The relative strengths and limitations of the various analysis approaches have been studied by academics with specific respect to algorithms; dynamic analysis methods are generally accepted to be much more powerful in conducting effective audits.[149]

3.14 Before inspecting the data and code, it is important to have the organisation's documentation, pseudo-code and general explanations. This helps an uninformed third-party to understand the context for the development and deployment of the algorithm. The documentation might include communications and internal documents about the business context, objectives, design, architectural diagrams, training (including relevant function(s) that has been maximised during an algorithm's training stage), key performance indicators (KPIs), and monitoring of algorithmic systems.[150] If the company runs its own audits, it might also be possible to ask for the output of these, be it fact sheets,[151] model cards,[152] transparency reports,[153] or other internal reviews.[154]

3.15 Next, analysing both input and output data (similarly to those outlined above) can provide much richer information. This occurred in the European Commission's investigation of Google's comparison-shopping service, which was found to have self-preferenced their rankings. For example, the

---

[148] Edward W. Felten & Joshua A. Kroll, SCI. AM. (Apr. 16, 2014), Heartbleed Shows Government Must Lead on Internet Security.
[149] Joshua A. Kroll , Joanna Huey , Solon Barocas , Edward W. Felten , Joel R. Reidenberg , David G. Robinson & Harlan Yu Accountable Algorithms, 165 U. Pa. L. Rev. 633 (2017), Accountable Algorithms
[150] In the CMA's Online Platforms and Digital Advertising market study, a principle was proposed around making high level "objective functions" of algorithms transparent in order to allow proportionate regulatory oversight. Appendix U: supporting evidence for the code of conduct – paras. 160-165.
[151] Arnold, M, Bellamy, R, Hind, M, Houde, S, Mehta, S, Mojsilovic, A, Nair, R, Ramamurthy, KN, Reimer, D. Olteanu, A, Piorkowski, D, Tsay, J, & Varshney, K (2019), 'FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity'.
[152] Mitchell, M, Wu, S, Zaldivar, A, Barnes, P, Vasserman, L, Hutchinson, B, Spitzer, E, Raji, ID, & Gebru, T (2019), 'Model Cards for Model Reporting', in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, January, pp. 220-229.
[153] Datta, A, Sen, S, & Zick, Y (2016), 'Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems', *IEEE Symposium on Security and Privacy*.
[154] Raji, ID, Smart, A, White, RN, Mitchell, M, Gebru, T, Hutchinson, B, Smith-Loud, J, Theron, D, & Barnes, P (2020), 'Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing', in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 33-44.

Commission analysed 1.7 billion search queries and simulated swapping the ranking of search results to see if the number of clicks changed, which formed an important part of their investigation.[155] If the algorithm is a machine learning algorithm, inspecting the data used to train the model could reveal bias or fairness issues, as the data itself is often the source of harm rather than the (often generic) optimisation methodology used.[156]

3.16 In the situation where it is possible to access the code, but not the full decision system/infrastructure or the data, it may be helpful to also look through the code of the algorithm itself, undertaking a comprehensive audit either through manual review or in an automated manner.[157] This was done by the Australian Competition and Consumer Commission (ACCC) in the case against Trivago, where experts were brought in to analyse both the code and the input and output data.[158] However, unless the algorithm is relatively simple, this can be a highly challenging task. Indeed, more complex machine learning algorithms are often called 'black box' algorithms because even the developers responsible for them do not necessarily fully understand how they map a given input to an output.[159] Moreover, some complex algorithms may only lead to harmful outcomes in the context of a particular dataset or application.[160] This approach is therefore likely of use only in specific cases and dependent on the nature of the algorithm.

3.17 If an external party wants to audit an algorithm and is given access, it may be made easier if that access is provided via an API. Given that companies often process data in a variety of ways (e.g. via HTTPS requests, or pre-existing APIs for third-party access) these APIs may possibly be made available to auditors with little extra effort. An alternative is a technological solution that uses a third-party sandbox in which algorithms can be shared by their owners and analysed in a privacy-preserving manner by appropriate external parties. Such approaches may allow external parties to carry out meaningful audits without needing to have access to the data or code; this can be of particular

---

[155] European Commission (2017), 'Statement by Commissioner Vestager on Commission decision to fine Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service', 27 June.
[156] Datta, A, Fredrikson, M, Ko, G, Mardziel, P, & Sen, S (2017), 'Proxy Discrimination in Data-Driven Systems: Theory and Experiments with Machine Learnt Programs', *arxiv preprint.*
[157] Sandvig et al. review this "Code Audit" methodology comparatively with other audit methodologies. Sandvig, C, Hamilton, K Karahalios, K & Langbort, C (2014), 'Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms', *Paper presented to "Data and Discrimination," a pre-conference of the 64th annual meeting of the International Communication Association, Seattle, WA, USA*.
[158] Federal Court of Australia (2020), 'Australian Competition and Consumer Commission v Trivago N.V. [2020] FCA 16'. 20 January.
[159] There is significant research into methods that facilitate explanation and comprehension of entire fields of complex algorithms – for example: Montavon, G, Samek, W, Müller, K-R (2018), 'Methods for interpreting and understanding deep neural networks', *Digital Signal Processing*, Vol. 73, pp. 1-15, ISSN 1051-2004.
[160] Sandvig, C, Hamilton, K Karahalios, K & Langbort, C (2014), 'Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms', *Paper presented to "Data and Discrimination," a pre-conference of the 64th annual meeting of the International Communication Association, Seattle, WA, USA*.

value where data is sensitive or encrypted.[161] There are proposed solutions in active development,[162] but they require further work before they can be widely deployed in practice.[163]

3.18 Alternatively, a randomised control trial (RCT) can be used to conduct an end-to-end audit, which might be the most effective method to audit harm for many algorithmic decision processes. We note that web-facing companies are frequently running large numbers of RCTs internally,[164] and therefore may be able to easily support such an RCT for audit, depending on its exact nature.[165] It could also be possible to perform auditing by repurposing pre-existing RCTs that have already been conducted.

3.19 As outlined, the potential efficacy of an algorithmic audit of the nature laid out in this section is greater than that of a "black-box" investigation of the sort outlined in Section 3.1. However, the degree to which such audits are successful can depend significantly on the willingness of the corporate owner of an algorithm to collaborate, share information and reduce friction in access to code and data. Therefore, the ability of a regulator to effectively audit algorithms in this manner is likely to heavily depend both on the provision of appropriate incentives for companies to positively engage, for example through relevant legislation or soft power, as well as the existence of effective formal information gathering powers.

4. The role of regulators in addressing these harms

4.1 The CMA has a mission to make markets work well for consumers, businesses, and the economy. For each of the theories of harm set out in section 2, there is a role for consumer and competition enforcement to address these harms. In addition, competition agencies, regulators and policymakers are working to adapt the scope of the law as well as their enforcement policies. As markets evolve, we will use our existing tools, and develop new capabilities and argue for greater powers where needed to prevent harms to competition and consumers as they arise.

---

[161] Royal Society (2019), *Protecting privacy in practice: The current use, development and limits of Privacy Enhancing Technologies in data analysis.*
[162] Epstein, Z, Payne, BH, Shen, JH, Dubey, A, Felbo, B, Groh, M, Obradovich, N, Cebrian, M, Rahwan, I (2018), 'Closing the AI Knowledge Gap', *arxiv preprint.*
[163] United Nations (no date), *UN Handbook on Privacy-Preserving Computation Techniques.*
[164] Exact figures are typically secretive, however, in a paper from 2013, Microsoft describe running over 200 concurrent experiments to over 100 million users of their Bing search engine. Kohavi, R, Deng, A, Frasca, B, Walker, T, Xu, Y, & Pohlman, N (2013), 'Online Controlled Experiments at Large Scale', in *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining.*
[165] An analysis of the use of A/B tests and their impact on individuals can be found in Jiang, S, Martin, J, & Wilson, C (2019), 'Who's the Guinea Pig? Investigating Online A/B/n Tests in-the-Wild', in *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 201-210).

4.2 There is a strong case for intervention:

(a) The opacity of algorithmic systems and the lack of operational transparency make it hard for consumers and customers to effectively discipline firms. Many of the practices we have outlined regarding online choice architecture are likely to become more subtle and challenging to detect.

(b) Some of the practices we outline involve the algorithmic systems of firms that occupy important strategic positions in the UK economy (and internationally).

4.3 Both these factors suggest that market forces are unlikely to effectively disrupt many of the practices we have outlined, with a potential for significant harm from exploiting consumers and excluding competitors.

4.4 In pursuing our mission, we will work together with other regulators such as the ICO and the EHRC where appropriate.

4.5 In this section, we set out some actions that the CMA and regulators more broadly can take. By setting out these ideas, we aim to promote useful discussion with potential stakeholders, and to contribute to building consensus, both in the UK and internationally, about the tools and powers that regulators will need in order to discharge their duties effectively.

4.1 Provide guidance to businesses and set or clarify standards

4.6 There have been a range of expert reports that have set out the potential benefits of clearer guidance and standards to help businesses to understand what is expected of them and how they can comply with relevant law.[166] For example, in a recent report, the Centre for Data Ethics and Innovation emphasised the importance of clear regulatory standards in realising the key benefits of AI.[167]

4.7 There have also been numerous reports and initiatives that set out advice and guidelines for AI,[168] and frameworks to help businesses to quality assure their

---

[166] See Royal Society and British Academy (2017), 'Data management and use: Governance in the 21st century'; Hall, W & Pesenti, J (2017), 'Growing the artificial intelligence industry in the UK'; and Bakewell, JD, Clement-Jones, TF, Giddens, A, Grender, RM, Hollick, CR, Holmes, C, Levene, PK et al. (2018) 'AI in the UK: ready, willing and able?' *House of Lords Select Committee on Artificial Intelligence.*

[167] For example, clear regulatory standards on how firms can access sensitive data to be able to train models in a fair way. See Centre for Data Ethics and Innovation (2020), 'CDEI AI Barometer', 18 June.

[168] See for example: European Commission High-Level Expert Group on AI (2019), 'Ethics Guidelines for Trustworthy AI', 8 April; Whittaker, M, Crawford, K, Dobbe, R, Fried, G, Kaziunas, E, Mathur, V, West, S M, Richardson, R, Schultz, J, Schwartz, O (2018), 'AI now report 2018'; The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2019), 'Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems'; Abrassart, C, Bengio, Y, Chicoisne, G, de Marcellis-Warin, N, Dilhac,

machine learning algorithms appropriately – particularly to guard against the concerns around bias set out in the section on 'Discrimination' above. For example, in an April 2020 blog, the US Federal Trade Commission stressed the importance of being transparent, explaining decisions to consumers, ensuring that inputs are accurate, and that inputs, processes, and outcomes are fair.[169] Algorithm risk assessments and impact evaluations have been proposed to help organisations to design and implement their algorithmic systems in an accountable, ethical and responsible way[170], while businesses themselves have begun to develop open-source tools to assess bias and fairness in their own algorithms.[171] Businesses could also support the development of assessments and tools by setting up internal or external ethical oversight mechanisms, and sharing best practices.[172] We support the development of ethical approaches, guidelines, tools and principles being developed in both the public and private sectors. However, we note that many of the guidelines and principles are not legally binding[173] and regulators and policymakers may need to go further.

4.8 One form of impact assessment that is mandated under data protection legislation is a Data Protection Impact Assessment (DPIA). Organisations must undertake a DPIA when they process personal data in AI systems to systematically analyse, identify and minimise data protection risks of a project or plan before they start, as a core part of their accountability obligations. In addition to guidance on how to complete a DPIA, the Information Commissioner's Office also produced a framework for auditing AI, which provides guidance on how to assess the risks to the rights and freedoms that AI can pose from a data protection perspective, including how to mitigate

---

M-A, Gambs, S, Gautrais, V, et al. (2018), 'Montréal declaration for responsible development of artificial intelligence'; OECD (2019), 'OECD Principles on AI'; Partnership on AI (2018), 'Tenets'; Future of Life Institute (2017), 'Asilomar AI principles'.

[169] US Federal Trade Commission (2020), 'Using Artificial Intelligence and Algorithms', 8 April. See also US Federal Trade Commission (2016), 'Big Data – A Tool for Inclusion or Exclusion?', January.

[170] See Ada Lovelace Institute & DataKind UK (2020), 'Examining the Black Box: Tools for Assessing Algorithmic Systems'; Raji, D, Smart, A et al. (2020), 'Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing', in *Conference on Fairness, Accountability, and Transparency*, p33–44. [online] Barcelona: ACM; and Reisman, D, Schultz, J, Crawford, K & Whittaker, M (2018), 'Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability', *AI Now*.

[171] Open source toolkits have been developed by several companies including LinkedIn (Vasudevan, S & Kenthapadi, K (2020), 'LiFT: A Scalable Framework for Measuring Fairness in ML Applications', in *Proceedings of the 29th ACM International Conference on Information & Knowledge Managemen*, pp. 2773-2780), IBM (Bellamy, RKE, Dey, K, Hind, M, Hoffman, SC, Houde, S, Kannan, K, Lohia, P, Martino, J, Mehta, S, Mojsilovic, A, Nagar, S (2019), 'AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias', *IBM Journal of Research and Development*, Vol. 63, Issue 4/5) and Google (Wexler, J, Pushkarna, M, Bolukbasi, T, Wattenberg, M, Viegas, F, & Wilson, J (2019), 'The what-if tool: Interactive probing of machine learning models', *IEEE transactions on visualisation and computer graphics* 26, 1, 56-65).

[172] European Commission Independent High-Level Expert Group on Artificial Intelligence (2019), 'Ethics Guidelines for Trustworthy AI'.

[173] Wagner, B (2018), 'Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?', in M. Hildebrandt (Ed.), *Being Profiling. Cogitas ergo sum*. Amsterdam University Press.

them. Both pieces of guidance help organisations to think through how to address some of the algorithmic harms we have outlined in this paper.

4.9 Another frequent theme of AI reports is encouraging firms to be transparent about their algorithmic systems.[174] Explaining how an algorithm works and how it produces a particular output is key to this. We think this is important for consumers and businesses who are affected by the decisions of these systems,[175] so that affected parties can themselves identify problems and potentially resolve these with firms without regulatory intervention, although we note that transparency alone is often not sufficient to fully address the issues. Explainable artificial intelligence (XAI) is a research field that offers various methods and approaches for explaining different AI methods, or parts of the AI-assisted decision-making process. Where firms are using simpler algorithms, such as logistic regression or a simple decision tree, explaining how an output was reached can be straightforward. With more complex 'black box' models, other methods can be used, for example proxy models, which can approximately match the system and produce an explanation. Visualisations, sensitivity analysis or saliency maps can also help understand which input features are most influential in producing an output.[176] The ICO and The Alan Turing Institute have produced guidance on explaining decisions made with AI, and emphasise the need to tailor the explanation to the person receiving the explanation.

4.10 Auditors or regulators will also require significantly detailed explanations of the outputs of firms' algorithmic systems, in order to allow them to undertake thorough technical inspections. In a July 2020 blog post, the then-CEO of TikTok stated that 'all companies should disclose their algorithms, moderation policies and data flows to regulators', and outlined some steps that TikTok had taken to do so. For more complex algorithms, this would need to include explainable algorithmic systems. We note that in some cases, extensive transparency and granular explanations should only be shared with those that need to understand them (i.e. auditors or regulators) given that they could increase the risk of unwanted manipulation of the algorithms (i.e. 'gaming') or, in the case of pricing algorithms, facilitate collusion by market participants.

4.11 Indeed, in preparation for potential regulatory intervention, we suggest it is incumbent upon companies to keep records explaining their algorithmic systems, including ensuring that more complex algorithms are explainable.

---

[174] For example, see European Commission (2020), 'Algorithmic Awareness-Building'.
[175] For example, there are transparency requirements on platforms in the EU platform-to-business (P2B) regulation. See European Commission, 'Platform-to-business trading practices'.
[176] See for example Guidotti, R, Monreale, A, Ruggieri, S, Turini, F, Giannotti, F, & Pedreschi, D (2018), 'A survey of methods for explaining black box models', *ACM computing surveys (CSUR)*, *51*(5), 1-42 and Molnar, C (2020), *Interpretable Machine Learning*.

We believe companies should be ready to be held responsible for the outcomes of their algorithms especially if they result in anti-competitive (or otherwise illegal or unethical) outcomes. The guidance produced by the ICO and The Alan Turing Institute outlines the policies, procedures and documentation that firms could provide to regulators.[177] Where appropriate, there should also be design documents that record system goals, assumptions about sociotechnical context, and other important considerations before development even commences.[178]

4.12 Regulators could also work with consumers, firms, developers and other experts to formulate implementable standards and guidance for good practices to reduce the risk of harm. With clearer standards and guidance, firms may have a stronger incentive to take steps to design and build transparency and accountability processes into their algorithmic systems, instead of leaving them as afterthoughts. The European Commission High-Level Expert Group on AI suggests that co-regulatory approaches beyond standards can be developed, such as accreditation schemes and professional codes of ethics. Firms that invest in sound data governance, monitoring and keeping records of the behaviour and decisions of their algorithmic systems are better able to identify and mitigate risks, and to demonstrate compliance when needed.

4.13 Stringent legal requirements to address risks already exist for investment firms engaged in algorithmic trading. The Commission Delegated Regulation 2017/589 sets out a number of technical and organisational requirements for investment firms, such as the need for formalised governance arrangements with clear lines of accountability, as well as monitoring mechanisms to address issues before and after trades. It also stipulates that a firm's compliance staff should have a basic understanding of how the algorithmic systems work, and that they should have continuous contact with technical staff responsible for the system's operation.[179] Additional legislation (including orders following a CMA market investigation) or guidance could clarify the extent to which firms outside of the investment sector have similar requirements for their algorithmic systems.

---

[177] An additional tool that could provide the basis for standards in record keeping for explainability is PROV-DM, which is a data model (knowledge graph) that provides a vocabulary for the provenance of data. It describes what a decision system does, as well as the people, organisations and data sets involved, and how data has been attributed and derived. See Huynh, TD, Stalla-Bourdillon, S, & Moreau, L (2019), 'Provenance-based Explanation for Automated Decisions: Final IAA Project Report'.

[178] For example, in Raji, I.E. et al. (2020), 'Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing', they introduce a framework to do this.

[179] See Articles 1 and 2 of European Commission (2016) 'Commission Delegated Regulation (EU) 2017/589'. *Official Journal of the European Union*, L87/417.

4.2 Identify and remedy existing harms

4.2.1 Intelligence gathering

4.14 The CMA's Data, Technology and Analytics (DaTA) Unit monitors complaints, the press, and research papers to identify potential leads on the algorithmic harms identified in this paper. The team may then use some of the approaches outlined in Section 3.1 to test potential theories of harm and identify potential cases proactively.

4.15 If a formal investigation is opened, the relevant information gathering powers allow us to request information to allow more in-depth analysis and auditing, using methods such as those outlined in Section 3.2. The Digital Markets Taskforce advises that the DMU should also be able to use information gathering powers to proactively monitor digital markets, as well as to pursue formal investigations.[180]

4.16 Other parties, such as researchers, investigative journalists, and other civil society organisations, frequently undertake their own investigations into potential harms. We invite interested parties to contact us with relevant leads and, where relevant, to collaborate with us on developing and applying methods to identify competition and consumer harms. Where appropriate, we may then take any investigation further using our information gathering powers.

4.17 We are also interested in the design and use of relevant software that can help consumers to protect themselves, for instance, by identifying situations where they are being presented with a personalised experience.

4.2.2 Formal investigations and remedies

4.18 Where we have reasonable grounds for suspecting that a business's use of algorithms may have infringed consumer or competition law, the CMA may open a formal investigation. Alternatively, the CMA can launch a market study or market investigation on any matter related to the acquisition or supply of goods and services in the UK which may have an adverse effect on the interests of consumers.

4.19 In these circumstances, we can use a range of powers to obtain information and data as discussed in the previous section on techniques (see Section 3.2). These information gathering powers cover data, code, and documentation, and include requirements on firms to help us to understand

[180] See Appendix G of the Advice of the Digital Markets Taskforce.

and test their algorithmic systems. Strong information gathering powers are essential to monitor and investigate algorithmic systems effectively, which is why we recommended that the DMU should have sufficient information gathering powers to do so.

4.20 Where we have found a problem, the CMA expects those responsible to put an end to problematic conduct and prevent it from arising in the future. Our intervention powers depend on which legal tool and approach we adopt to address these harms, and whether we are running a competition case, consumer enforcement, or a market study. Firms can offer remedies to the problem, which we may accept if we think that it will address our concerns. Alternatively, in some circumstances, we may design and impose remedies on firms. In either case, we test proposed remedies so that they are effective, proportionate, and (where relevant) we can monitor compliance and evaluate their effectiveness. Depending on our findings and legal approach, our remedies may be limited to one or a few firms, or they may be market-wide. We also recommended that the DMU should have strong remedial powers.

4.21 Remedies are context-specific, depending on the nature and scale of the problems we find. To illustrate what we might do:

(a) We may order firms to disclose information about their algorithmic systems to consumers.

(b) We may require a firm to disclose more detailed information to approved researchers, auditors and regulators, and to cooperate with testing and inspections. Cooperation may involve providing secure access to actual user data, access to documentation and internal communications on the design and maintenance of the algorithmic system, and access to developers and users for interviews. These audits may be ad hoc or more regular.

(c) We may impose ongoing monitoring requirements and require firms to submit compliance reports, provide ongoing and continuous reporting data or API access to key systems to auditors and regulators.

(d) We may require firms to conduct and publish algorithmic risk assessments of prospective algorithmic systems and changes, and/or impact evaluations of their existing systems.

(e) We may order firms to make certain changes to the design and operation of key algorithmic systems and require firms to appoint a monitoring trustee to ensure compliance and that the necessary changes are made.

4.22 Pursuing formal investigations has wider strategic benefits. Formal cases can clarify the law and provide an opportunity to apply and sharpen relevant policy. They can provide an example to others, potentially deterring other firms from harming consumers. In addition, the experience and learning from formal investigations will feed back into better guidance and tools (discussed above) to help firms to comply.

4.3 Ongoing algorithmic monitoring

4.23 Algorithmic systems are often updated regularly with new or evolving datasets and dynamic models, meaning that one-off audits may become quickly outdated. It will therefore be important for algorithmic systems that could bring about competition and consumer harms on an ongoing basis to be monitored, with penalties for violations where they occur.

4.24 Monitoring could take several forms. The tools for gathering intelligence outlined above (4.2.1) could be used to check up on algorithmic systems that have been investigated in the past, including press reports, and research undertaken by academics and civil society organisations.

4.25 Regulatory sandboxes[181] can be useful for regulators to provide a safe space for firms to technically test algorithms in a live environment, without being subject to the usual regulatory consequences.[182] This was recommended as a potential approach for the DMU in the Digital Markets Taskforce advice. The FCA and ICO have used regulatory sandboxes to allow experimentation, data sharing, and to provide a secure environment for using sensitive information to detect and mitigate biases.[183] The use of sandboxes may be particularly fruitful when regulators possess sensitive data necessary to audit algorithms that firms do not have themselves.[184] However, sandbox testing is challenging to implement. It can be difficult to design a sandbox environment which has external validity and adequately represents the environment in which the algorithm will operate. It could also be difficult to ensure that the algorithm does not exhibit different behaviour under test conditions than when it has been deployed for some time. This is a particular concern where the behaviour of the algorithm depends on its interactions with other businesses'

---

[181] A regulatory sandbox is a programme run over a set number of months, in which firms can test their products with real customers in a controlled environment under the regulator's supervision and feedback, whilst not being subject to the usual rules that apply to regulated firms. A key aim is to assess the viability of innovations in terms of their compliance with regulatory requirements. See UNSGSA (2020), 'Early Lessons on Regulatory Innovations to Enable Inclusive FinTech: Innovation Offices, Regulatory Sandboxes, and RegTech'.
[182] Financial Conduct Authority (2015), 'Regulatory sandbox'.
[183] Information Commissioner's Office (2020), 'Blog: ICO regulatory sandbox'.
[184] Veale and Binns discuss a similar solution to mitigating discrimination without access to protected characteristics data via what they call "trusted third parties" tin Veale, M, & Binns, R (2017), 'Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data', *Big Data & Society*, *4*(2).

conduct (which may also be driven by other algorithms), or where the algorithm is expected to change over time as it learns. Sandbox testing may therefore need to be repeated at regular intervals.

4.26 More consistent and regular monitoring could occur if businesses are required to employ compliance staff and report on compliance with remedies we require, or standards that we might set to tackle the harms outlined above.[185] This might also help facilitate the growth of a market for specialised firms to provide algorithmic auditing services to prove certification against these standards.[186] Regulators can go beyond this; for example, the Financial Conduct Authority has a robust supervisory function that undertakes proactive investigations into possible harms, focusing on those businesses that pose the greatest harm.[187]

4.27 In the UK and internationally, there have been numerous reports recommending the creation of a new regulatory function or body for ex ante regulation of digital platforms.[188] The Government has now announced the establishment of a Digital Markets Unit within the CMA from April 2021. The Digital Markets Taskforce has recommended the DMU have powers to enforce a code of conduct for strategic market status (SMS) firms and to impose pro-competitive interventions. It has also recommended that the DMU have information gathering powers to monitor digital markets more widely, beyond the SMS regime. In each case, these powers will need to be sufficient to suspend, block, and reverse conduct which is enabled or facilitated through algorithmic practices.[189]

4.4 Build and use digital capabilities, and enhance collaboration

4.28 The CMA has invested in wider data and technology skills, with specific investments related to algorithms. We have recruited data scientists and engineers, technologists, and behavioural scientists as part of a Data, Technology and Analytics (DaTA) team, to develop and deploy new analytical and investigative techniques, and to broaden our range of evidence and

---

[185] For example, these could build on the IEEE Ethically Aligned Design Initiative to develop certification that an algorithmic system is transparent, accountable and fair.
[186] Although we do acknowledge such certification processes may be particularly challenging for algorithms, which can be constantly evolving, subject to frequent updates, and can exist as part of a broader system which itself may change.
[187] Financial Conduct Authority (2019), 'FCA Mission: Approach to Supervision'.
[188] These include: 'Unlocking digital competition, Report of the Digital Competition Expert Panel', March 2019; Crémer et al. (2019), 'Competition policy for the digital era'; ACCC (2019), 'Digital platforms inquiry - final report'; Stigler Centre Committee on Digital Platforms (2019), 'Final Report'; and the CMA (2020), 'Online platforms and digital advertising market study final report'.
[189] Considerations will have to be made about whether tests or audits should be carried out or verified by independent third parties to ensure that they are conducted properly; repeated regularly to capture changes in the algorithm; and whether a precautionary principle would need to be applied in assessing the risk of harm from an algorithm prior to deployment.

intelligence. We have used our new capabilities to monitor businesses and markets, to gather and pursue potential leads, to assist our conduct of formal investigations, and to design and implement effective remedies. Other regulators may also benefit from investing in similar capabilities, and we welcome the opportunity to share our experience.

4.29 Subject to any legal restrictions, we can also collaborate with other regulators by sharing information, such as complaints submitted that might indicate where an algorithmic harm is arising, or where there are specific cases that raise issues for multiple regulators, such as data protection and consumer issues. In particular, we will continue to collaborate with the ICO and Ofcom and develop our joint capabilities as part of the Digital Regulation Cooperation Forum (DRCF).

4.30 We are also undertaking more international collaboration, including cooperation between competition authorities.[190] This cooperation facilitates mutual learning, sharing of good practices and insights, and the possibility for the sharing of code and tools, where these do not contain any confidential information. Where key algorithms and systems are applied globally, we intend to continue working with our counterparts to build international consensus on effective regulation and standards.

5. Conclusions

5.1 Algorithms are an integral part of how many markets and firms operate. They have greatly enhanced efficiency and allowed firms to deliver better products and services to consumers. However, firms may also misuse them, whether intentionally or unintentionally, and can cause harms to consumers and competition, often by exacerbating or taking greater advantage of existing problems and weaknesses in markets and consumers. In this paper, we have categorised and discussed a number of potential harms to consumers and competition arising from the misuse of algorithms. We have focused on areas that are particularly relevant to the CMA's mission to make markets work well for consumers, businesses and the economy.

5.2 Whilst there has been a lot of attention and discussion of algorithmic harms in general, there is relatively little empirical work on some of the specific areas of consumer and competition harms, and almost none that we are aware of in the UK. In particular, we found gaps in work surrounding the operation and effects of automated pricing on collusion, techniques to efficiently identify and

[190] For example, we have established a subgroup of data science groups under the auspices of the Multilateral Mutual Assistance and Cooperation Framework for Competition Authorities.

assess the impact of personalisation, and the same for manipulative choice architecture more generally.

5.3 Of course, cases and literature will generally reflect the current (or historical) situation. But digital markets are dynamic. Deployed algorithmic systems and the data that they use can change quickly as techniques improve, so new harms may manifest quickly. For example, even assuming that algorithmic collusion is not a significant problem now, it could rapidly become so if a critical mass of firms starts to use more complex algorithmic systems for pricing in a particular market (for example by adopting third party solutions).

5.4 Firms maximise profit. In pursuing this objective, without adequate governance, firms designing machine learning systems to achieve this will continually refine and optimise for this using whatever data is useful. Algorithmic systems can interact with pre-existing sources of market failure, such as market power and consumers' behavioural biases. This means that using some algorithmic systems may result in products that are harmful. As regulators, we need to ensure that firms have incentives to adopt appropriate standards and checks and balances.

5.5 The market positions of the largest gateway platforms are substantial and appear to be durable, so unintended harms from their algorithmic systems can have large impacts on other firms that are reliant on the gateway platforms for their business. If algorithmic systems are not explainable and transparent, it may also make it increasingly difficult for regulators to challenge ineffective measures to counter harms.

5.6 Due to the various harms identified in this paper, firms must ensure that they are able to explain how their algorithmic systems work.

5.7 This paper accompanies the launch of the CMA's analysing algorithms programme. We will work with others to identify problematic markets and firms violating consumer or competition law, take cases forward where action is required. We will work with other regulators and industry to set standards and determine how algorithms should be audited.