Appendix I: search quality and economies of scale

Introduction

1. This appendix presents some detailed evidence concerning how general search engines compete on quality. It also examines network effects, scale economies and other barriers to search engines competing effectively on quality and other dimensions of competition.

2. This appendix draws on academic literature, submissions and internal documents from market participants.

Comparing quality across search engines

3. As set out in Chapter 3, the main way that search engines compete for consumers is through various dimensions of quality, including relevance of results, privacy and trust, and social purpose or rewards.[1] Of these dimensions, the relevance of search results is generally considered to be the most important.

4. In this section, we present consumer research and other evidence regarding how search engines compare on the quality and relevance of search results.

Evidence on quality from consumer research

5. We reviewed a range of consumer research produced or commissioned by Google or Bing, to understand how consumers perceive the quality of these and other search engines. The following evidence concerns consumers' overall quality judgements and preferences for particular search engines:

6. Google submitted the results of its Information Satisfaction tests. Google's Information Satisfaction tests are quality comparisons that it carries out in the normal course of its business. In these tests, human raters score unbranded search results from mobile devices for a random set of queries from Google's search logs.

   - Google submitted the results of 98 tests from the US between 2017 and 2020. Google outscored Bing in each test; the gap stood at 8.1 percentage points in the most recent test and 7.3 percentage points on average over the period.

[1] As discussed in Appendix H, search engines also compete over access to consumers, through the default search positions on web-browsers and devices.

- Google submitted the results of 25 tests from the UK between February and December 2019. Google outscored Bing in each test; the gap stood at 10 percentage points in the most recent test and was 8.5 percentage points on average over the period.

7. Google explained that removing all Wikipedia results from the Google Search results would cause Google's Information Satisfaction number to drop by 0.6 points. Thus, Google said that a 7 percentage point gap between Google and Bing is equivalent to the effect of losing the Wikipedia database approximately 12 times over. Google said that the Information Satisfaction scores measure mobile search quality, although Google has no reason to believe that the gap in Information Satisfaction scores between Google and Bing would be materially different on desktop.

8. Google also submitted a Comparative Quality Analysis (dated July 2019), in which users of mobile devices were asked to rate Google's results side-by-side alongside those of Bing, DuckDuckGo, Qwant and Ecosia. The document concluded that 'Google quality is much better than others in all locales tested' (which included the US and the UK). As part of the Comparative Quality Analysis, 200 UK consumers compared Google results to those of another search engine in response to 'random queries'. In this test, Google scored more highly than each competitor, and Bing performed relatively better against Google than DuckDuckGo and Qwant did. These results are summarised below:

   - Google vs. Bing comparison: for 21% of queries users preferred Google's results, for 1.5% of queries users preferred Bing's results, and for 77% of queries users judged the results to be about the same.

   - Google vs. DuckDuckGo comparison: for 37% of queries users preferred Google's results, for 0% of queries users preferred DuckDuckGo's results, and for 64% of queries users judged the results to be about the same.

   - Google vs. Qwant results comparison: for 44% of queries users preferred Google's results, for 0.2% of queries users preferred Qwant's results, and for 56% of queries users judged the results to be about the same.

9. Microsoft submitted results from its Bing Challenge consumer research in which desktop users compare Google and Bing results side-by-side. In the most recent period (Q4 19):

   - When brands were visible, 42% preferred Google, 23% preferred Bing, and the remainder were draws.

- When brands were hidden, 37% preferred Google, 35% preferred Bing and the remainder were draws.

10. Google generally scored more highly than Bing in the studies above, although the studies are not entirely consistent, in that Google scored more highly than Bing in all Google studies, whereas Google and Bing received a similar number of preferences in the unbranded version of Microsoft's Bing Challenge. In both Google's Comparative Quality Analysis and Microsoft's Bing Challenge, there were a significant number of instances in which users did not express a strong preference for either Google or Bing's results.

11. The studies above provide some evidence of the perceived quality gap between Google and Bing being smaller when tests are undertaken on an unbranded basis. This could suggest that objective quality differences are limited for some queries and/or that consumers find it difficult to compare search results.

12. Some of the other evidence that we have reviewed suggest some consumers struggle to tell the difference between search engines in the absence of brands and logos. For example, Google submitted the results of a survey of [✂] users, who were asked to identify search engines based on their Search Engine Results Page where logos had been removed. When a Google results page was shown, [90-100%] of users correctly identified it as Google. However when a Bing or Yahoo! results page was shown, nearly half wrongly identified it as Google.

13. Some of the consumer research that we reviewed provided insights into the type of search quality differences that exist between search engines.

14. Microsoft submitted consumer research on 'switching queries' (dated 2019), which shows the extent to which users of the Internet Explorer and Edge browsers use Google Search rather than Bing for particular types of query. Queries where regular Bing users were relatively less likely to use Google included [✂]. Queries where regular Bing users were more likely to use Google included [certain queries including some with a location-based element]. The document discusses three examples of 'high switch' queries where a relatively large proportion of users tried Google after entering the same query in Bing: [✂].

15. As discussed above, Google's Comparative Quality Analysis found a narrower gap in search result quality between Google and Bing than between Google and search engines that use syndicated Bing results, such as Ecosia and DuckDuckGo. The research highlights several potential reasons for this, including:

- Result localisation: the report indicates that, unlike Bing, DuckDuckGo does not return localised results in response to queries without an explicit location.[2] The document says that DuckDuckGo's 'imprecise location yields less relevant results'.

- Search features: the document says that 'Google leads in search features (coverage and utility)'. It also indicates that Bing has a greater range of features than providers that syndicate Bing's search results. It states that 'Ecosia's quality is much worse than Bing' and that Ecosia 'lacks features that Bing provides'.

16. Microsoft submitted the results of qualitative research involving discussions with a small number of Bing users in the UK. This indicated that the perceived relevance of results was a source of relative weakness for Bing in the UK. [✂]

17. In Microsoft's Bing Challenge research, consumers were invited to select reasons that explained their preference for Google or Bing's results. Many reasons were selected a similar number of times. We note that popular responses including those that suggest a general preference for a particular brand[3] (in the branded version of the test), those that relate to different aspects of relevance[4] and those that indicate a preference for more images and videos[5]. [✂]

18. The results of the first three studies above (Microsoft's switching queries study, Google's Comparative Quality Analysis and Microsoft's qualitative research) are consistent with the idea that search engine quality comparisons vary depending on the type of query. These studies also provide some evidence that quality differences between Google and other general search engines are more pronounced in relation to some local-based queries and certain other categories of query, including those where search features and [✂] play a role in answering the query.

19. Microsoft told us that Google's perceived advantage especially applies to uncommon queries (also known as 'tail queries') and that Google has richer local and specialty results.

---

[2] We note that DuckDuckGo summitted that it does return localized results in response to queries without an explicit location, but DuckDuckGo does so by using a representative location derived from the IP address, and not a more precise location. DuckDuckGo users can obtain localized results based on a more precise location, provided the user expressly opts into that method (which is still anonymous). It noted that the opt-in for localized results using a more precise location is not yet available in the UK within the DuckDuckGo Privacy Browser Android app. See https://help.duckduckgo.com/duckduckgo-help-pages/privacy/anonymous-localized-results/ for more details.

[3] [✂].

[4] [✂].

[5] [✂].

20. The studies that we reviewed did not categorise queries according to whether these were common (ie head queries) or rare (ie tail queries). Therefore, the consumer research that we reviewed does not provide direct evidence as to whether quality differences may vary between head and tail queries. We discuss tail queries further in the section below on ‘scale effects in click-and-query data’.

Other evidence on search quality

21. We also reviewed some other evidence and submissions suggesting that Google’s English-language search results are generally perceived as being of a higher quality than those of alternative general search providers. For example:

- Microsoft also submitted some qualitative research that indicated that the perceived relevance of results was a source of relative weakness for Bing in the UK, and an internal memo stating that Bing was 'trailing' Google on relevance in a number of regions outside of the US.[6]
- [a search engine] said that Google has the ‘best algorithmic results’.
- Apple said that it selected Google as the default English-language search engine for UK users ‘because it is the best English Language search engine’ and said that this decision was validated by [✂].[7]

22. On the other hand, in its response to our interim report, The Horizon Digital Economy Research Unit questioned the impartiality of previous tests conducted by Microsoft and suggested that ‘the dominance of Google in the general search engine market is not necessarily an indication of its overall better performance’.[8]

23. We also heard that search engines that access organic links through syndication agreements may not produce results of equal quality to those of the provider of those links. Consistent with the consumer research evidence, Ecosia noted that ‘the search results that we receive from Microsoft are not of the same quality as they show on Bing’. [✂]

24. Various academic studies have sought to rank the performance of general search engines. One method of analysis pursued by academics is to select a

[6] [✂].
[7] As discussed in Appendix H, internal documents we viewed suggest that search quality is an important factor in the way that search defaults are selected, alongside the level of compensation offered by the search engine to the device manufacturer and other factors.
[8] Horizon’s response to our consultation on our interim report.

small number of queries, often between 10-20, and analyse a search engines' effectiveness based on pre-defined metrics such as the precision of results or relative recall.[9] Most of these studies note that it is difficult to objectively assess the relevance of search results since human assessment, which is prone to biases, is necessary to judge relevance.

25. These studies generally find that Google out-performs or matches other search engines in effectiveness. Shafi & Salib (2019), for example, found that Google obtained the highest precision and relative recall of 15 physical science queries, followed by Yahoo and Bing.[10] Further research by Goutam & Dwivedi (2012) found that on the basis of user effort measures,[11] Google was a more effective search engine than Yahoo or MSN.[12]

26. It should be noted that these reviews are based on a very small sample of queries whereas search engines handle a vast number of queries every day, with queries ranging in complexity and popularity. The results of these studies should therefore be interpreted with caution. A study by Lewandowski (2015) investigated Google's and Bing's effectiveness in German utilising a sample of 1,000 informational queries and 1,000 navigational queries.[13,14] Lewandowski found that Google outperforms Bing in both query types although the difference in performance was much lower for informational queries.

27. We also reviewed some research concerning the drivers of search quality. Lerner (2014) posits that engineering talent combined with crawling and indexing technology are vital inputs to search quality and goes on to state that Google has developed superior web crawling and indexing capabilities.[15] Lerner highlights a blog post by a Microsoft engineer which explained that a 'viable' search engine requires 'hiring enough smart people' and that 'the Bing search relevance staff is a fraction the size of Google's.'[16] McAfee (2015) similarly presented that because of Google's greater scale it can improve the

---

[9] Relative recall is the ability of a search engine to obtain all or most of the relevant documents in question. As it is not possible to catalogue all documents on the internet, academic researchers typically measure relative recall according to the sum of all documents obtained by all search engines being analysed.
[10] Shafi, S.M. & Sabah, A. (2019). *Retrieval performance of select search engines in the field of physical sciences.*
[11] Search engines were measured according to ranked precision, which took into account relevance of results, session duration, dwell time, ranked precision, clicks hits and user satisfaction of results.
[12] Goutam, R. & Dwivedi, S. (2012). *Performance evaluation of search engines via user effort measures.*
[13] Informational queries refer to queries where the user seeks information on a topic they are interested in while navigational queries refer to where users aim to navigate to an already-known web page.
[14] Lewandowski, D. (2015). *Evaluating the retrieval effectiveness of web search engines using a representative query sample.*
[15] It is noted that Lerner received assistance from Google's team during this research. Lerner, A.V. (2014). *The role of 'Big Data' in online platform competition.*
[16] Clemmer, A. (2014). *What 'viable search engine competition' really looks like*, available at http://blog.nullspace.io/building-search-engines.html [last accessed: 25/06/2019].

quality of search pages for rare queries much faster than other search engines.[17]

28. In the section below, we consider barriers to competition on search quality.

Barriers to competition on search quality

Click-and-query data

29. When a consumer enters a query, the search engine uses algorithms to select the most relevant result from its index. Search engines collect and store aggregated click-and-query datasets containing information about what consumers searched for and how they interacted with the results that they were served.

30. Click-and-query data helps search engines to understand how well their product is performing and to identify and test potential improvements, such as changes to ranking and spelling correction algorithms. It is primarily used to enhance the quality of the search algorithm, as results that consumers have clicked on previously in response to the same or similar queries are a useful signal of the relevance of a webpage. Useful interaction data may include, for example, dwell time on the search results page, whether the consumer scrolled down the page, whether the mouse hovered over a particular element, what links the consumer clicked and whether they hit the back button after following a link.

31. Other uses of click-and-query data include the development of natural language processing models to better understand the meaning of a query and suggest corrections to common spelling mistakes, as well as the construction of logical relationships among queries to suggest related searches.

32. In our interim report, we said that the greater scale of English-language queries seen by Google is likely to support its ability to deliver more relevant search results compared to its competitors, especially in relation to uncommon and fresh queries. In turn, we said that, lack of comparable scale in click-and-query data is likely to limit the ability of other search engines to compete with Google.

[17] It is noted that McAfee was a chief economist at Microsoft. McAfee, P. (2015). *Measuring scale economies in search*.

33. In reaching this position, we considered a range of evidence including submissions from search engines and other parties, as well as academic literature.

34. Our review of academic literature found that increased quantities of user data, and click-and-query data in particular, can improve the quality of ranking although this is dependent upon the extent to which the results are personalised. In explaining this finding we can distinguish between two cases:

- Changes in ranking quality not related to result personalisation. While there is relatively limited empirical evidence that investigates the issue, the literature that exists appears unanimous in claiming that additional click-and-query data at query level improves the quality of ranking.[18] The size of this effect, however, varies greatly depending on the scale of the data held for individual queries. Indeed, the empirical evidence finds rapidly diminishing returns to scale, hence making this effect more relevant for rare queries (which make up a substantial share of the queries submitted to search engines).[19]

- Changes in ranking quality related to result personalisation. The empirical evidence on this point is split, with some authors finding a positive effect and others finding no effect of larger quantities of click-and-query data at individual-user level on the quality of ranking.[20]

35. Parties' submissions also highlighted a positive relationship between query scale and ability to return relevant results, with the strength of the relationship varying depending on the type of query. For example, views from parties included:

- DuckDuckGo said that 'more searches on DuckDuckGo means that we can more easily conduct anonymous A/B testing because we can more quickly obtain the needed sample sizes, and conduct more tests simultaneously'.

- Mojeek said that 'scale of queries is helpful for things like spelling suggestions and tail queries (need billions of queries to do spelling well)'.

- Microsoft said that 'the algorithms that power search perform better with more data and volume'. It said that there are 'diminishing marginal returns

[18] See Schaefer, M, Sapi, G and Lorincz, S (2018); He, D, Kannan, A, Liu, TY, McAfee, RP, Qin, T and Rao, JM (2017), 'Scale Effects in Web Search', International Conference on Web and Internet Economics pp294-310.
[19] See He, D, Kannan, A, Liu, TY, McAfee, RP, Qin, T and Rao, JM (2017).
[20] See Schaefer, M, Sapi, G and Lorincz, S (2018); Chiou, L and Tucker, C (2017), 'Search engines and data retention: Implications for privacy and antitrust', National Bureau of Economic Research.

to additional scale once a search platform obtains a certain size' but that scale effects are important for 'tail queries' and 'fresh queries'.

- Google said that it derives 'little marginal benefit from collecting additional click and query data about head queries' and that there 'may be value to having a greater amount of click and query data for "tail queries" (queries entered infrequently by users)'. It said that there are often more efficient approaches to improving the results for tail queries than increasing scale of data. For example, implementing a misspelling algorithm.

36. In responding to the position that we expressed in our interim report, Google argued that concerns regarding click-and-query data were 'overstated' and that 'access to click and query data is not a barrier to entry', whereas other search engines were supportive of our assessment of this issue, for example, Verizon Media said that 'greater [click-and-query] data scale enables greater relevance which is a key aspect of quality for consumers'.[21]

37. Since the publication of the interim report, we have refined our understanding of the size and distribution of search engines' click-and-query datasets and have considered further arguments and evidence in relation to the link between data scale and search quality.

Additional analysis of the size and distribution of query data seen by Google and Bing

38. Since the publication of our interim report, we have calculated shares of supply on the basis of the volume of searches provided by the parties. As set out in Appendix C, this analysis shows that on average Google's share of the total volume of searches undertaken on search engines in UK in 2019 was 93% and Bing's share was 5%. In absolute terms, around [150-200 billion] search events took place on Google and [0-10 billion] took place on Bing in the UK in 2019.

39. We asked Google and Microsoft about the internal definitions that they use to define queries according to frequency. Both parties indicated that they do not use precise, formal definitions.

- Google said that:

[21] Verizon Media's response to our consultation on the Interim Report.

(i) ‘Head’ queries are those queries that are seen more than five times per day and account for 15% of search queries in a given period (eg a month, six months, or a year).

(ii) ‘Mid’ or ‘torso’ queries are those entered relatively frequently. These are typically seen twice per day. Google also considers as ‘torso’ queries those queries that are more frequent than tail queries, but less frequent than head queries.

(iii) For working purposes, ‘tail’ or ‘longtail’ queries are those that are contained in a query set when adding together the least common queries until they amount to 25% of total traffic.

- Microsoft said that:

(i) ‘Head’ queries are the most common queries representing the top roughly 20% of the query frequencies.

(ii) ‘Body’ queries represent the next approximately 50% of the query frequencies.

(iii) ‘Tail’ queries account for the last 30% of query frequencies.

40. Google submitted that having a larger dataset does not necessarily mean seeing ‘tail’ queries more often. In support of this, it submitted the results of its own analysis of the number of ‘singleton’ (ie unique) queries that were present in Google’s US English-language dataset during a one-day period. Google started by considering a small subset of the total queries that it observed. As larger samples were considered [✂].

Analysis of one-week of query data seen by Google and Bing

41. In order to better understand the query data seen by Google and Bing, we requested and analysed data for all the search events seen by Google and Bing in the UK within a one-week period.[22]

42. In the analysis that follows, each and every search that is undertaken is counted as a ‘search event’. The text that is associated with a search event is a ‘query’. Queries that appear only once are ‘singletons’. The set of queries that remain once duplicates have been removed are ‘distinct queries’. For

[22] The query data for the two search providers covered slightly different time periods. The Google data was collected on Pacific Time (PT) basis, and started/ended 8 hours later than the Bing data which was collected on Greenwich Mean Time (GMT) basis. We do not expect this to affect our results substantially.

illustrative purposes, we define the head as the 15% of queries seen most often in a dataset and the tail as the 30% of queries seen least often.

- *Overview of datasets*

43. The Google dataset contained [3-4 billion] search events, compared with Bing's [0.1-0.2 billion], giving a combined total of [3-4 billion] search events.[23]

44. As shown in Table I.1, the Google dataset contained [3-4 billion] non-blank, non-spam search events, compared with Bing's [0.1-0.2 billion]. For the remainder of this analysis, we focus exclusively on non-blank, non-spam queries – even when not explicitly stated.

Table I.1: Number of search events seen by Google and Bing

Engine	*Number of (Non-Blank) Search Events*
Bing	[3-4 billion]
Google	[0.1-0.2 billion]

Source: CMA analysis of Google and Bing data

45. Below, we first discuss the number of distinct queries that were observed, then the overlap between the Google and Bing datasets, and then the shape of the data distributions.

46. As shown in Table I.2, the Google dataset contains [0.8-1.0 billion] distinct queries compared to Bing's [50-100 million], making the number of distinct queries in the Google dataset approximately 16 times larger than the Bing dataset.

Table I.2: Number of distinct queries seen by Google and Bing

Engine	*Number of (Non-Blank) Distinct Queries*
Google	[0.8-1.0 billion]
Bing	[50-100 million]

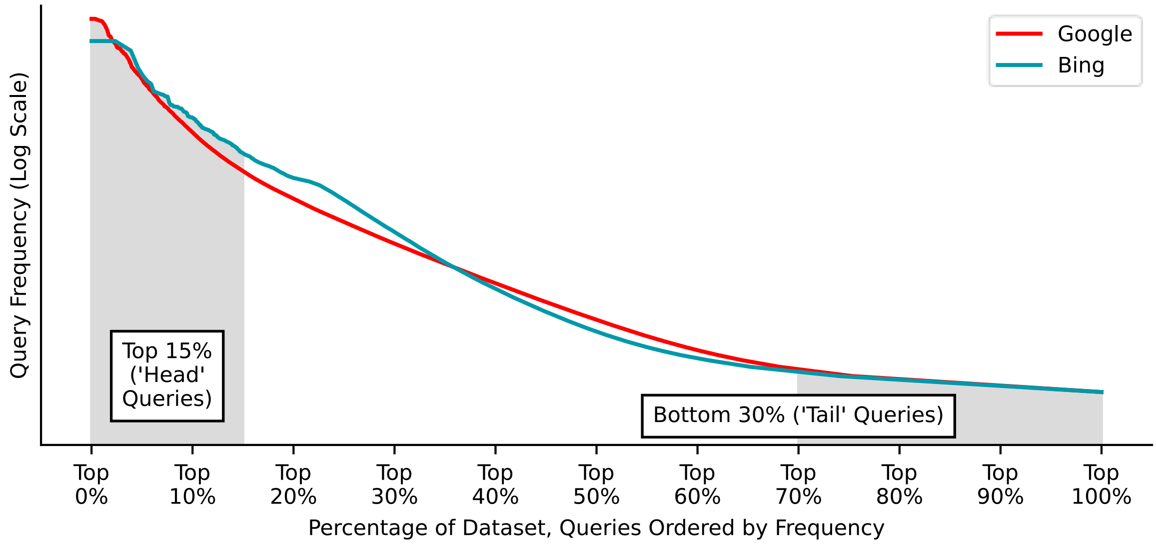
Source: CMA analysis of Google and Bing data

47. In both datasets, the distribution of the number of search events per distinct query (ie the query frequency) was highly skewed, with a small group of 'head' queries appearing frequently, and a large number of uncommon queries appearing less than 100 times in the week we examined.

[23] Blank queries make up [0-5]% of the Google dataset, and [0-5]% of the Bing dataset. Queries denoted as 'spam' by Google also make up [10-20]% of the Google dataset, while the Bing data submission already excluded spam queries.

48. Figure I.1 shows, for Google and Bing, the query frequency as a function of the cumulative percentage (ie top 10%, top 20% etc) of search queries in the dataset ordered by frequency. Due to the high skewness of the distributions, the query frequency is presented in a logarithmic scale in powers of 10.

Figure I.1: Cumulative frequency distribution of non-blank, non-spam search events in the Google and Bing search datasets, by cumulative percentage of dataset.



Source: CMA analysis of Google and Bing data

- *Overlap between datasets*

49. Table I.3 shows how much of the combined Google and Bing dataset overlaps and how much is exclusive to one or other of the search engines. It shows that 50% of all search events in the combined dataset are for queries that appear exclusively in the Google dataset, 2% of events are for queries that appear exclusively in the Bing dataset, and 48% of events are for queries that are seen in both datasets. When it comes to distinct queries, 94% appear exclusively in the Google dataset, 4% are exclusively in the Bing dataset, and 2% are in both datasets.[24]

[24] It should be noted the data submitted to us by Google and Bing was for raw queries ie before the correction of any spelling mistakes. Had we seen queries after the correction of spelling mistakes, we expect that the number of distinct queries would have reduced significantly and that the degree of overlap between the datasets would likely have increased.

Table I.3: Proportion of the total number of search events and district queries in the combined dataset that only appeared in one of the search engines' datasets

	Search Event Share (Total)	*Search Event Share Proportion (Total) (%)*	*Distinct Query Share*	*Distinct Query Share Proportion (%)*
Total	[✂]	100	[✂]	100
Google Exclusive	[✂]	50.2	[✂]	93.9
Bing Exclusive	[✂]	1.7	[✂]	4.1
Shared	[✂]	48.1	[✂]	2.0

Source: CMA analysis of Google and Bing data

50. Of the distinct queries in the combined dataset, nearly all were seen only by Google, and very few were seen by both search engines, or only by Bing.

51. We note that the overlap between Google and Bing is significantly greater on a 'search event' basis than on a 'distinct queries' basis; a large part of the overlap on a 'search event' basis is likely to relate to a small number of common queries.

52. Table I.4 shows an alternative expression of the overlap between the Google and Bing datasets. It shows that:

- Approximately 47% of all search events in the Google dataset are for queries that were also searched in the Bing dataset, compared with 69% of events in the Bing dataset which also appear in the Google dataset.

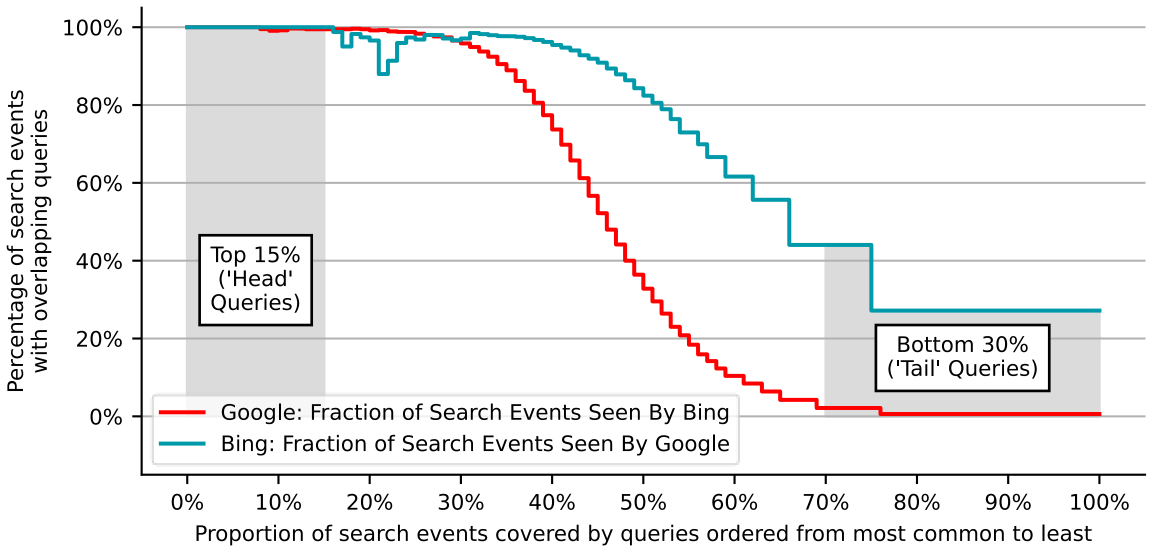
- Approximately 2% of the distinct queries in the Google dataset appear in the Bing dataset, while 34% of the distinct queries in the Bing dataset appear in the Google dataset.

Table I.4: Proportion of search events and district queries in one search engine's dataset that also appeared in the other search engine's dataset

Engine	*Total Number of Shared (Non-Blank, Non-Spam) Search Events*	*Proportion of Shared (Non-Blank, Non-Spam) Search Events (%)*	*Number of Distinct Shared (Non-Blank, Non-Spam) Queries*	*Proportion of Distinct Shared (Non-Blank, Non-Spam) Queries (%)*
Google	[✂]	46.8	[✂]	2.1
Bing	[✂]	69.3	[✂]	33.5

Source: CMA analysis of Google and Bing data

Figure I.2: Distribution of the percentage of Google non-blank, non-spam search events which were for queries seen by Bing, and vice versa, by the frequency of their search query



Source: CMA analysis of Google and Bing data
Notes: The 0%-1% percentile on the x-axis represents the most common queries (ie the top of the head) and the 99%-100% percentile represents the least common queries (ie the bottom of the tail).

53. The red line in Figure I.2 shows the percentage of Google search events that were for queries also seen by Bing, broken down by percentile of search events ordered by query frequency. The queries most seen by Google (0%-1% percentile) are on the left and those least seen (99-100% percentile) are on the right. Similarly, the blue line shows the percentage of Bing search events that were for queries also seen by Google, broken down by percentile of search events ordered by query frequency.

54. Figure I.2 shows that the fraction of search events relating to queries that are also seen by the other search engine changes depending on how common or rare the query is, with ‘head’ queries being much more likely to be seen by both engines than ‘tail’ queries. The proportion of Google’s head queries seen by Bing, and of Bing’s head queries seen by Google, is close to 100%. This means that the most common queries in the two engines tend to be seen by both. For a small section of the Bing distribution, around the 25th percentile, the proportion of search events for queries that are also seen by Google is slightly lower than the proportion of Google search events that are also seen by Bing.[25] From around the 30th percentile to the rest of the distribution, Google sees more queries for Bing’s search events than the other way around.

[25] We have not been able to interpret this result as we did not have access to the raw text of users' queries (we only had access to a hashed version that is not human interpretable).

55. We also undertook some more detailed analysis of the respective head and tail portions of the Google and Bing datasets.

- *Further analysis of tail queries*

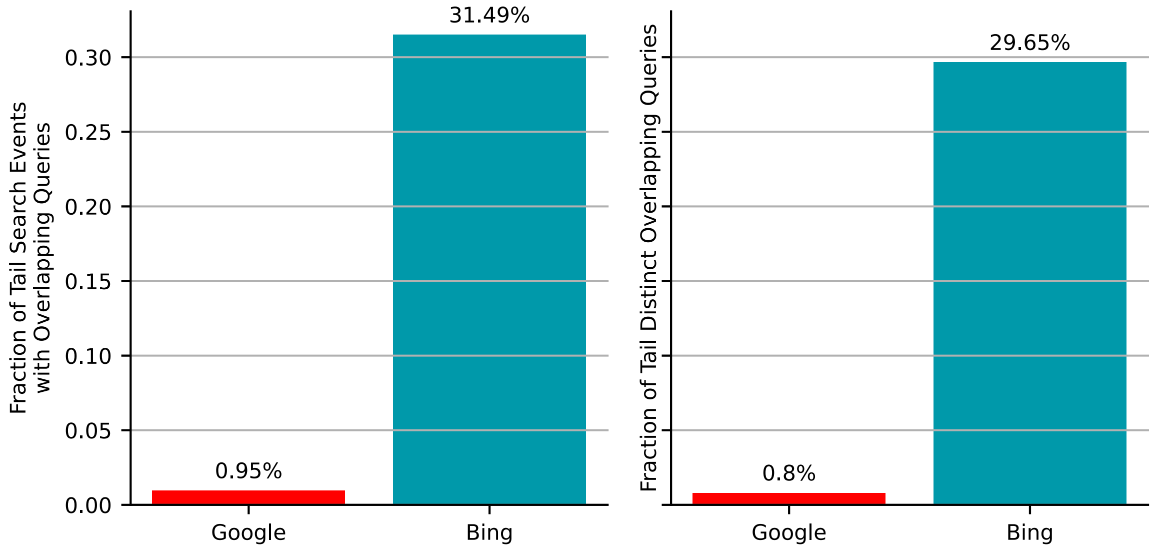
56. We define 'tail' queries as queries which appear a total of once or twice in the weeks' worth of data that we examined for each engine.

57. Search events that are for a query Google only saw once or twice made up [30%-40%] of the search events in its dataset. Search events for a query Bing only saw once or twice made up [30%-40%] of the search events in its dataset. Search events for a query that only appeared once in their dataset (otherwise known as 'singletons') made up [20%-30%] of the Google dataset and [20%-30%] of the Bing dataset.

58. For distinct queries, those which were seen once or twice made up [80%-90%] of the distinct queries in the Google dataset and [80%-90%] of the distinct queries in the Bing dataset. Singletons made up [70%-80%] of the distinct queries in the Google dataset and [70%-80%] of the distinct queries in the Bing dataset.

59. Figure I.3 shows the fraction of 'tail' search events (left) and distinct queries (right) where the query is seen by the other search engine. It shows that around 1% of Google 'tail' search events are for queries which are seen by Bing. In comparison, 31% of Bing 'tail' search events are for queries which are seen by Google. For distinct queries, 0.8% of Google's 'tail' distinct queries are seen by Bing, whereas 30% of Bing's 'tail' distinct queries are seen by Google.

Figure I.3: Fraction of tail search events (left) and distinct queries (right) in which the query is seen in the other dataset.



Source: CMA analysis of Google and Bing data

Additional submissions from Google

60. In its submissions since our interim report, Google repeated its position that there are diminishing returns to scale from additional data.

61. In support of this, Google said that the Microsoft-Yahoo (syndication) deal of 2010 'doubled Bing's query volume overnight but failed to improve the relevance or monetisation of Bing's search queries'. In relation to this example, the news reports that we have seen indicate that the Microsoft-Yahoo deal fell short of Yahoo's expectations in terms of the revenue per search earned by Yahoo. However, these reports do not contain estimates of how the deal impacted the relevance of Bing's search results.[26,27] We note that the Microsoft-Yahoo agreement would have increased the query volume seen by Bing, while still leaving it with a significantly smaller scale than Google.

62. Google also submitted an example based around the search results served by Google and Bing in response to the query 'Japan tsunami' following the natural disaster in 2011. According to Google's submission, Google Search was returning relevant results 24 minutes after the disaster, whereas Bing was still showing irrelevant results 650 hours later 'despite having (presumably) received a large number of Tsunami-related queries by that

[26] The Wall Street Journal, 'Yahoo Wants Out of Microsoft Deal', 7 May 2013. Accessed 8.4.2020.
[27] Searchengineland.com, 'As The Yahoo-Microsoft Search Alliance Falls Short, Could A Yahoo-Google Deal Emerge?, 19 July 2012. Accessed 8.4.2020.

time'. In Google's view, this confirms that smaller rivals' access to more query data does not necessarily lead them to improve the relevance of their search results. We have not seen detailed evidence in relation to this example. Even if Google did return more relevant results than Bing for a given number of searches for this query, this would not necessarily indicate that data scale is not important. For example, the greater scale of data seen by Google prior to the disaster may have supported its ability to develop and train the algorithms and other technology that determined the search results that it served.

63. Google submitted that major improvements in the relevance of search results have instead come from technological and analytical developments that do not depend on having more data. For example, Google said that its BERT technology for understanding the intent behind queries[28] and its Query Deserves Freshness model for identifying and promoting 'fresh' web pages have made large improvements to Google's search results and did not depend on large-scale access to uncommon queries.[29]

64. We note that search technology has developed in various ways over the last decade and that the role of click-and-query data scale, more general cost-based economies of scale, and other factors such as engineering expertise is likely to have varied depending on the innovation in question.

65. In some cases, the development of technology improvements may not have been contingent on large click-and-query datasets. For example, we understand that Google's BERT technology for understanding query intent was initially developed using Wikipedia text as training data. However, even where new technologies are initially developed without using click-and-query data, our understanding is that being able to test and refine algorithm updates across a large volume of search queries may support the deployment of these technologies. We note that some of the technology improvements highlighted by Google relate to the way that Google indexes web-pages. For example, the Caffeine update and Query Deserves Freshness model. We discuss search indexing in the following section.

[28] https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html
[29] Google submitted that Bidirectional Encoder Representations from Transformers (BERT) enables Google's search engine to understand the context of words in a search query and how they fit together, rather than looking at words in isolation. Therefore, Google now recognizes that the query "2019 3 brazil traveler to usa need a visa" refers to Brazilian visitors to the US, not the reverse. With this better understanding of language, Google can show more relevant results. Another example is Google's Query Deserves Freshness model, which looks at 'spikes' in the dates when pages are first indexed by Google to identify 'fresh' results.

Our assessment

66. The evidence that we have reviewed shows that:

- **There are advantages to scale in click-and-query data** – search engines that see more queries (and more consumer responses to those queries) can engage in increased experimentation and learning about what consumers want and have greater possibilities to iterate and improve their service.

- **The marginal benefit of additional data depends on the type of query** – where a search engine sees a search query very frequently (sometimes referred to as 'head queries'), then the marginal benefit from seeing that query more often is relatively lower. Conversely the marginal benefit of seeing a query more often is higher for uncommon queries (sometimes referred to as 'tail queries').

67. Google – as the most used search engine – sees each uncommon query more times than Bing in a given time period. We consider that, in particular, this supports Google's ability to serve more relevant results to uncommon queries compared to Bing. Even if click-and-query data only helped Google return more relevant results for a modest proportion of search queries, this would further reinforce consumers' perceptions of Google as the highest-quality search engine and make them less inclined to consider alternative providers.

68. Overall, our assessment is that the greater scale of English-language queries seen by Google supports its ability to deliver more relevant search results compared to its competitors. We consider that this effect is more material for particular types of query, such as uncommon or 'tail' queries. Given the importance of search relevance to consumers, the lack of comparable scale in click-and-query data limits the ability of other search engines to compete with Google.

Search indexing

69. In order to return relevant results in response to a range of consumer queries, search engine algorithms must be able to draw on indices that cover a very wide range of relevant webpages, and that provide an up-to-date picture of the content of those webpages.

70. To crawl the web for new or updated webpages, search engines follow links from other known webpages and use data on known webpages' URL addresses and the links on webpages. Search engines also make use of crawl requests and sitemaps submitted by webmasters (ie people who are responsible for maintaining websites).

71. Search engines record and organise data and metadata collected from crawlers on the content of webpages to form an index, from which relevant search results are drawn. This data can include the title of a webpage, the words it contains and their location within the webpage, as well as metadata on the author of the page and the time the page was last updated.

72. We heard that search engines face two potential barriers to developing a competitive web-index: cost-based economies of scale and network effects that impact the ability of web-crawlers to access and index webpages.

73. In the sections below, we review submissions and evidence regarding Google and Bing's web-indices and barriers to developing a competitive search engine.

Google and Bing's indices

74. Google and Microsoft's Bing are the only two search engine providers that maintain at-scale English-language web indexes. We therefore requested and reviewed submissions and internal documents regarding how Google and Bing's web-indices compare and the reasons for this.

75. We found that Google's index is larger than that of Bing in terms of number of pages in the index. Based on submissions from these parties, Google's index contains around [500-600 billion] pages and Microsoft's index contains around [100-200 billion] pages.

76. Google submitted that the effectiveness and efficiency of indexing is more important than the size of a web index. Google, therefore, attempts to optimise its index to find relevant information as quickly as possible. Google said that, with the launch of its 'Caffeine' update in 2011, it learned to analyse the web in smaller portions and update its index on a real-time basis (ie as soon as Google's web crawlers cover a webpage, Google's index is updated instantly for that individual webpage).

77. Microsoft noted that to build a competitive index they must be able to discover, select, crawl and rank the best URLs. Microsoft's internal documents additionally listed five metrics upon which they evaluate their index selection which are as follows: comprehensiveness, efficiency, latency, quality and ranking accuracy,

78. Google said that its crawlers are superior at seeking-out relevant search results to tail queries. [✂][30]. Google said that it moved to mobile-first indexing in 2019 (having experimented with this since 2016), whereas Bing has declined to do so [✂].

79. Internal documents submitted by Microsoft indicate that, in 2019, Bing's index was on a par with Google for frequently visited URLs but less comprehensive than Google's for URLs that were visited less frequently.

80. Microsoft submitted that overall, it considers its index competitive with Google's. However it said that it believes Google would have an easier time in web crawling for three separate reasons:

- Google's position of strength across the web provides it many different sources of sensors for indexing.[31]

- Webmasters prioritise and provide significantly more signal to Google in order to ensure they are discoverable in the Google index.

- Google's web crawler is more welcome than Bing's crawler which makes it easier to discover new sites or updates.

81. We assess below the role of cost-based scale economies and the extent to which web-crawlers have different levels of access to web pages.

Economies of scale

82. The costs associated with crawling and indexing do not increase proportionally with the number of users of the search engine. Therefore developing a web-index is subject to economies of scale.

83. Crawling and indexing the web represents a significant cost for those search engines that do it. Microsoft estimated that its indexing investments added up to billions over time, while other estimates have suggested that Google and Bing spend hundreds of millions of dollars a year on this activity.[32]

---

[30] Google explained that it invested in analysing the web in smaller portions and updating its index on a real-time basis to improve the efficiency of its web crawler.
[31] Bing noted Google's sources of signals to include: Google Analytics, Android, Google Fiber, Google DNS Service, Google Tag Manager, Google Ads, YouTube and others.
[32] For example, the European Commission quoted DuckDuckGo as follows: 'Bing and Google each spend hundreds of millions of dollars a year crawling and indexing the deep Web. It costs so much that even big companies like Yahoo and Ask are giving up general crawling and indexing. Therefore, it seems silly to compete on crawling and, besides, we do not have the money to do so'. Source: Google Search (Shopping) Commission Decision (non-confidential version), 27 June 2017, page 66.

84. Google submitted that developing a good indexing and web-crawling infrastructure may require significant investment, but is a cost that Google has to face in the same way as its rivals and therefore does not represent a barrier to entry or expansion. It said that, if anything, indexing and crawling costs are lower for more recent entrants, since they can use free open source crawlers as a base (Google noted that Common Crawl's February 2020 archive alone contains 2.6 billion web pages).

85. Other web-crawling search engines include Yandex, Cliqz and Mojeek:[33]

- Yandex said that, while it does crawl English web pages, it has not prioritised building an at-scale English language index. This is due to the high sunk cost involved, limited access to distribution points, economies of scale, and issues with monetization of the search engine, rather than crawler-blocking. It would not make sense to invest in an at-scale English language index, unless it had an expectation of achieving a return on investment, which would in turn require access to distribution points and the ability to monetise effectively.

- Cliqz told us that it had invested [a significant double-digit million euro investment] into developing its own service. Cliqz suggested that it would require a 5-10% market share in order to attract advertisers, monetise effectively and earn a return on investment. However, this is not possible due to Google's default positions, which act as a barrier to it distributing its search services.

- Mojeek told us that its index contained around 2.5 billion pages, from an investment of two million pounds. It said that it was seeking to grow its index in a more cost-efficient way compared to other smaller players (such as Qwant which invested 30 million euros and Blekko which invested 60 million dollars). Mojeek said it is not looking to compete head-on with Google at present and does not need to in order to create a profitable company.

86. DuckDuckGo (which crawls the web and maintains partial web-indexes, but does not maintain its own complete web-index) said that it believed that the Common Crawl index could in theory be used by a search engine but in practice would not be a viable option. This is because, if Common Crawl is similar to the Internet Archive, the index would be not up-to-date and not comprehensive. DuckDuckGo also notes that, equally importantly, the Common Crawl index is not a viable option because it does not

[33] Yandex is a Russian multinational that has a more-than 40% share of search in Russia. Cliqz is a Germany-based search engine. Mojeek is a small UK-based search engine.

generate/create ranking signals (eg word frequencies) that interpret the crawl data.

87. As part of our analysis of Google's profitability (see Appendix D), we estimated the replacement cost that a competitor would incur if it sought to replicate Google's search engine, based on a review of data provided by Google, Microsoft and other search engines. On this basis we estimated that it could cost between £7.5 to £22.5 billion (equivalent of $10 to $30 billion) to create the technology to develop a search engine of comparable scale to Google, and to fund operating and development costs prior to reaching sufficient scale to be profitable. These figures include the cost of developing a web-index and also other costs such as compute and storage and networking costs.

Crawler access to web pages

88. Some parties raised crawler access to web pages as a barrier for smaller search engines developing web-indices, and we undertook further work to investigate this.

89. The mechanism for crawler-blocking is that webmasters (ie people who are responsible for maintaining websites) place robots.txt[34] files on their websites, requesting that some or all crawlers do not access all or parts of the website. We heard that website owners may have legitimate motivations for doing so. For example, motivations can include fraud prevention and avoiding the increased running costs that can result from a large number of automated bots crawling a website.

90. We heard that, when web-crawlers encounter blocking, search engine providers can contact webmasters to seek a change of policy. However, the effort and cost of doing so means that search engines that are subject to fewer blocking instructions may have an advantage.

91. Relatedly, webmasters often want their websites to be found on search engines, and it is likely that webmasters will prioritise submitting crawl requests and notifying updates to leading search engines, such as Google, and may neglect to do so for smaller search engines.

92. Microsoft said that a small fraction of sites have robot.txt files that enable Google to crawl the site but prohibit Bing. Microsoft gave several examples of websites that allow crawling by Google but prohibit crawling by Bing over all

[34] See further: description of robots.txt files.

or part of the site. These include or have included eBay's UK website,[35] the UK Passport Service website,[36] and the London Stock Exchange website.[37] Its main concern was that user impression of Bing search quality would be harmed if several high profile websites were not properly searchable.

93. Yandex and Mojeek indicated that crawler-blocking was not the main barrier to them expanding their English-language web-indices. Rather, economies of scale and other issues were more important. However, Cliqz said that it had incurred 'significant business development expenses' over the last 5 years by having to contact popular publishers to gain permission to crawl their websites.

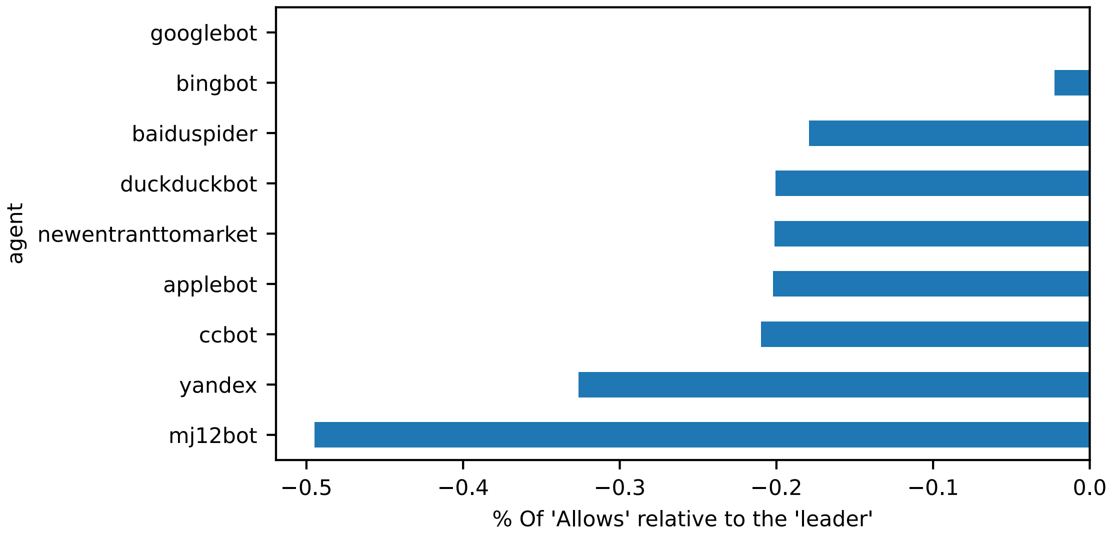
94. To test the significance of webmasters blocking the crawlers of smaller or new entrant search engines, we analysed a sample of 57 million domains from Common Crawl and found that 60% of sites in the sample hosted a robots.txt.[38,39] This means that almost two thirds of sites in the sample are implicitly blocking or limiting crawling.

95. Figure I.4 below shows the relative access to sites for different bots as the percentage less than the 'leader', the bot with 'access' to the most sites (ie Googlebot).

---

[35] eBay.co.uk robot.txt, accessed on 13 November 2019.
[36] UK Passport Service robot.txt, accessed on 15 October 2019.
[37] London Stock Exchange robot.txt, accessed on 15 October 2019.
[38] The data for this sample were taken from Common Crawl's July 2019 archive of robots.txt. Common Crawl describe themselves as a non-for-profit dedicated to providing a copy of the internet to researchers, companies and individuals at no cost for research and analysis. Their monthly samples are intended to be representative of the internet.
[39] We set a rule to each robots.txt such that a bot is (i) 'allowed' if it had complete access to every part of a website (ii) 'denied' if it had no access to any part of a website and (iii) 'denied partially' if it had access to some parts but not others. A bot is considered as having 'access' to a domain if it is either (i) allowed or (iii) partially denied access to the domain. The following bots were compared: (i) Googlebot: Crawler for Google's searchable index, (ii) Bingbot: Crawler for Microsoft Bing search engine (iii) DuckDuckBot: DuckDuckGo's web crawler (iv) Yandex: Crawler for Yandex, a Russian search engine; (v) Baiduspider: Crawler for Baidu, a Chinese search engine; (vi) Applebot: Crawler for Apple; (vii) MJ12Bot: Crawler for a UK based specialist search engine; (viii) CCBot: Crawler for Common Crawl, the source of this dataset; and (ix) NewEntrantToMarket: a fictitious crawler, used to assess how a new crawler would behave at the time the robots.txt were collected

Figure I.4: Relative access to sites for different bots



Source: CMA

96. The figure shows that Google's bot has the greatest access, followed by Bing's, with DuckDuckGo and new entrants more frequently denied access to sites. The effect is relatively small, with entrants having access to approximately 0.2% fewer sites than Google and Bing. The difference between Google and Bing is smaller still, with Bingbot having access to 0.02% fewer sites than Googlebot.

97. Therefore, differences in access levels provided to search engines appear to be small in numerical terms. We note that, even if a small number of popular sites cannot be crawled by some search engines, this could limit the ability of those search engines to return high-quality search results to certain queries.

Our assessment

98. Google and Microsoft are the only two providers that undertake English-language web-crawling and indexing at a scale that can support a competitive search engine in the UK. We consider that this reflects substantial scale economies in crawling and indexing, plus uncertainty for other search engines as to whether they can secure the other inputs (including sufficient scale in search queries and adverts) needed to earn a return on these investments. While network effects in indexing (and crawler blocking) appear to be a less significant issue than cost-based scale economies, in combination, these factors represent a barrier to entry and expansion for English-language web-crawling search engines.

Other supply-side barriers

99. We heard that some further supply-side barriers may limit the ability of search engines to compete on relevance of results.

Cost-based scale economies in search features

100. As set out in Chapter 3, search features, are an important part of how search engines' respond to search queries. For example, DuckDuckGo said that, as well as organic links, a competitive provider of general search services must have a set of high-quality search features including comprising: maps; local business answers (eg restaurant addresses and phone numbers); news; images; videos; products/shopping; definitions; Wikipedia reference; and quick answers (calculator, conversions, etc).

101. As set out earlier in this appendix, there is some evidence to suggest that the different search features that Google, Bing and syndicator search engines use explain some of the quality differences that exist between search engines.

102. Ecosia suggested that developing additional search content or features around the organic links that it syndicates from Bing is subject to economies of scale, because the product development effort is fixed, regardless of the number of consumers served.

103. On the other hand, Google suggested that any scale economies are limited: it said 'apart from limited economies of user data scale, Search does not enjoy significant benefits from its scale in other areas'.

104. As with web-indexing, we note that smaller search engines seeking to develop competitive features would likely face higher unit costs than larger search engines. In some cases, smaller players appear to have responded by obtaining search features and content from third parties. For example, DuckDuckGo uses Wikipedia for answers, Yelp for businesses, and Apple for maps.

Barriers to accessing location data

105. We heard that location data is an important input to producing relevant results. Location data can support the delivery of relevant results in at least two ways:

- Firstly, having access to search queries with a location signal helps to improve search algorithms, as search engines can learn how best to factor location into the ranking of results. We heard that mobile devices are a richer source of location data than desktop devices, and that data and learning from mobile devices can improve mobile and desktop search results alike. For example, Microsoft said 'mobile queries…help improve search algorithms with an extra dimension (location) that helps improve the engine's results in both mobile and PC search alike'.

- Secondly, having access to location data can help search engines to develop additional content and features, such as databases of local reviews, in order to supplement the organic website links that they collect from web-crawling. For example, we note that consumer research conducted by Microsoft (and discussed earlier in this appendix) suggested that Google has an advantage in [certain local queries] 'from Android phone location tracking, allowing it to track popular times and prompt users to submit reviews'.

106. Therefore, search engines without at-scale access to location data may be less able to provide relevant local search results.

107. Microsoft suggested that accessing at-scale location data from user devices is a critical input to providing relevant, localised results. It indicated its belief that Google has unique advantages in this area, due to the location data that it receives from the Android operating system and the location data it receives when users access Google Search or other apps like Google Maps/Waze.

108. We note that Google's high share of supply in general search, and especially in mobile search, suggests that it has access to many more queries with a location signal than other search engines. As set out in Appendix F, Google also has a significant advantage in relation to location data more generally. It gathers this data systematically and to a great level of detail from mobile devices running Android. As set out in Appendix G, on Android phones, half to two thirds of users have location services activated. Google also gathers location data through other apps like Google Maps/Waze which are popular also on non-Android devices.

Our assessment

109. Overall, we consider that cost-based scale economies in search features and Google's greater access to location data may provide Google with an advantage in relation to location-based queries and act as an additional barrier to expansion for other providers.

Conclusions

110. Competition over quality is an important mode of competition between search engines; quality is important for consumers and is one of the factors that

device manufacturers and other access point owners consider when choosing which search engines to set as default.[40] Relevance of search results is generally considered to be the most important element of search engine quality.

111. The evidence that we reviewed suggested that Google's search results are generally perceived to be of higher quality than those of Bing, although the research on this topic is not entirely consistent.

112. The key inputs to achieving relevant results include click-and-query data and an extensive and up-to-date web-index.

113. Click-and-query data is subject to scale effects. Google's greater scale supports its ability to iterate and improve quicker than other search engines and maintain a lead on search relevance. In contrast, rival search engines receive fewer queries and clicks, making it more challenging for them to improve the relevance of their search results.

114. Web-indexing is subject to cost-based scale economies. The costs of developing and maintaining an at-scale index are substantial and the ability of smaller search engines to repay these costs and earn a return on investment is contingent on their ability to secure the other inputs necessary to compete effectively in search. For example, they would also need to achieve scale in both search queries and search advertising, in order to offer relevant results and monetise effectively. Networks effects related to web-indexing and crawling-blocking provide a further advantage for Google relative to other search engines.

115. The development of search features, which sit alongside organic links on the search results page, is also subject to cost-based scale economies. Google's greater access to location data may provide it with an advantage in relation to location-based queries and act as an additional barrier to expansion for other providers.

116. Cumulatively, the factors above provide Google with a substantial advantage by supporting its ability to deliver higher quality results than its rivals. In turn, they act as a barrier to expansion for Bing and other potential competitors.

[40] See Appendix H for further information regarding competition for defaults.