IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

State of Colorado, *et al.*,

Plaintiffs,

v.

Google LLC,

Defendant.

Case No. 1:20-cv-03715-APM

HON. AMIT P. MEHTA

[redacted]

DEFENDANT GOOGLE LLC'S MOTION TO PARTIALLY EXCLUDE THE OPINION OF PLAINTIFFS' EXPERT JONATHAN BAKER

Pursuant to Federal Rule of Evidence 702, Defendant Google LLC ("Google") respectfully moves to partially exclude the opinions of Colorado Plaintiffs' proffered expert Jonathan Baker as set forth below, and any reliance thereon.

INTRODUCTION

As he readily admits, Professor Jonathan Baker is not an expert in any computer engineering field. Yet, among the opinions Colorado Plaintiffs ("Plaintiffs") seek to have him present are those involving search engine scale, an area which requires deep technical understanding of how search engines work. Professor Baker likewise strays from his training and experience in offering opinions on data privacy, an area in which he has no expertise. Unsurprisingly, given his lack of expertise, Professor Baker's "opinions" in these areas are primarily those of a narrator, selectively presenting record evidence from actual search engineers. And he otherwise seeks to bootstrap an "expert" opinion based on an interview with a third-party search engineer who was not permitted under the terms of the operative Protective Order in this case to review the relevant materials. Professor Baker's opinions in these two areas should be excluded.

Scale. Professor Baker offers opinions regarding the impact of the availability of user search data (or "scale") on search engine quality and competition among search engines. Relying on documents and deposition testimony, he asserts that greater scale benefits a search company by improving search and search advertising quality. And to support Plaintiffs' view of consumer harm, Professor Baker suggests that these materials show that any decrease in scale to Google will lead to less harm to Google's quality than the benefit it will provide to search competitors' quality (and presumably consumers will therefore be better off on the whole). But these opinions are highly dependent on an understanding of search engineering and how particular search engines

operate—including how search algorithms use various user data for testing and improving search results, and whether increased volumes of such data would meaningfully affect those technical processes. Professor Baker admits that he does not have any relevant experience with *search engine* scale, relying only on general concepts of economies of scale across all industries. Professor Baker's opinions on search engine scale are therefore not based on his knowledge or experience as Rule 702 requires and instead consist entirely of improperly repeating facts offered by others with actual technical experience.

Professor Baker also offers a rebuttal to Professor Edward Fox, a Professor of Computer Science at Virginia Tech retained as an expert by Google, who uses a data reduction experiment (DRE) to measure the effect on Google search quality resulting from a diminution in the amount of data available to train Google algorithms. But Professor Baker acknowledges that he does not have technical experience to rebut Professor Fox himself. Instead, Professor Baker relies solely on produced documents and [redacted]. Professor Baker cannot bootstrap an "expert" opinion through such an unreliable and unhelpful process in which Professor Baker himself apparently did not conduct any analysis. His opinions on scale should therefore be excluded.

Privacy. Professor Baker also opines regarding Google's and other search companies' privacy offerings. In particular, he asserts that record evidence suggests that search competitor DuckDuckGo has superior privacy protections and that greater competition would lead to "improved privacy protections." Ex. 1, June 6, 2022 Expert Report of Jonathan B. Baker at 18, 26. But, again, Professor Baker is neither a search engine engineer nor a data privacy expert and has no experience in the inherent tradeoffs between data use and other aspects of search quality, or any other aspect of privacy. He therefore offers nothing more than selectively repeating certain

documents and testimony, for which he has no specialized experience to assist the fact finder's understanding.

Because Plaintiffs have not carried their burden to establish that Professor Baker's opinions on scale and privacy are based on specialized knowledge that will help the trier of fact, his testimony on these topics should be excluded.

BACKGROUND

A. Fact Discovery on Scale and Privacy

Fact discovery in this case on the issues of scale and privacy was extensive. Documents regarding these topics were sought and produced, and the parties deposed a number of search engineers from various companies, including Google, Microsoft, DuckDuckGo, Apple, and Neeva, regarding these issues.

On scale, those witnesses testified regarding the various ways user data is utilized, the amount of user data needed to serve useful results and conduct experiments, and how technological advances have greatly decreased the amount of user data needed to offer quality search results and compete effectively. [redacted]

[redacted]

On privacy, documents and testimony from search engineers and others knowledgeable about search engine privacy discussed the trade-offs between data privacy and the quality and efficacy of search engines and search advertising products. Google search engineers testified about how Google "rel[ies] on data collection and data tracking to provide quality search results to users," including the technical "trade-offs [ ] in the kind of search experience you can get based on the data we have available to us." Ex. 8, Excerpts of the Deposition of Benedict Gomes, Ph.D, former Senior Vice President of Search at Google, (Dec. 10, 2021) at 246:13–247:6; *see also* Ex. 9, Excerpts of the Deposition of Catherine Edwards, Vice President of Engineering at Google (May 4, 2022) at 206:4–9 ("I am aware that we have had ongoing considerations of how much data we should log. Balancing the fact that many users view that data as valuable with the fact that some users view that as not valuable."). [redacted]

B. The Parties Have Engaged Other Experts on Scale and Privacy

In addition to fact discovery on these issues, the parties have retained other experts on the issues of scale and privacy. Google has retained:

- Professor Ophir Frieder, a professor of Computer Science at Georgetown University, to "assess, from a technical perspective, the principal elements of search quality, and whether there are factors other than large volumes of query data (beyond that already accessible to other search engines) that are important determinants of overall search quality." Ex. 11, Excerpts of June 3, 2022 Expert Report of Ophir Frieder at 3, 6.
- Professor Edward Fox, a professor in the Computer Science Department at Virginia Polytechnic Institute and State University, "to test the extent to which Google's search quality is affected by the volume of user interaction data available to train its ranking algorithms." Ex. 12, Excerpts of June 3, 2022 Expert Report of Edward A. Fox at 1, 5.
- Professor Catherine Tucker, a Professor of Management Science at MIT with an academic focus on "digital economics, and especially questions of how the increasing use of digital technologies has transformed advertising and questions of privacy," to assess various topics regarding search engine privacy. Ex. 13, Excerpts of June 6, 2022 Expert Report of Catherine Tucker at 1–2.

Colorado Plaintiffs have not retained their own technical experts on these issues. However, DOJ Plaintiffs have retained:

- Professor Douglas Oard, a professor at the College of Information Studies and the Institute for Advanced Computer Studies (UMIACS) at the University of Maryland, College Park, to provide a rebuttal to Professor Frieder and Professor Fox.
- Professor Kirsten Martin, a professor of Technology Ethics and of IT, Analytics, and Operations at the University of Notre Dame, to provide a rebuttal to Professor Tucker.

C. Professor Baker's Opinions

Professor Baker is an economist by training. He holds himself out as "an expert [ ] in economics, not on computing technology." Ex. 4, Excerpts of the Deposition of Jonathan Baker (Nov. 21–22, 2022) at 46:24–47:2. Professor Baker's Opening Report discloses his opinions regarding scale and privacy.

***Scale*.** Professor Baker opines that greater search query scale benefits a search engine in two ways: (1) search result quality and (2) the ability to serve search advertising. Regarding search result quality, he opines that "[g]reater scale, in terms of the number of users and search queries, ultimately helps a general search firm improve the quality of the search results it provides in response to specific search queries." Ex. 1 at 5. He claims "[t]hat is because a general search firm learns about the information users making a query are looking for—what counts as a good response—by evaluating user response to the choices offered by a SERP." *Id.* at 62. Regarding advertising, Professor Baker opines that "advertisers value search advertising more when the search firm has more search users." *Id.* at 5. And he claims that "greater user scale benefits advertisers by allowing general search firms to improve the match between advertisers and search users, thereby increasing the value of general search advertising to advertisers." *Id.* at 63. He further offers an opinion on the marginal effects of scale on Google's and other search competitors' quality and advertising, stating that the "increase in scale attributable to [the challenged] agreements would be unlikely to have generated more substantial improvements in Google's search and advertising quality as the same increase in scale would generate for Google's smaller rivals." *Id*. at 173.

In his reply report, Professor Baker discloses an opinion seeking to rebut Professor Fox's data reduction experiment (DRE).[1] He opines that Professor Fox's study is misleading in that it does not consider how "Google's search algorithm has benefitted from a long series of improvements over many years, each one of which benefited from Google's scale advantage at the time." Ex. 3, September 26, 2022 Expert Reply Report of Jonathan B. Baker at 64. And he asserts that the result of the experiment is [redacted]

Privacy. Professor Baker opines that "[i]n a more competitive general search services market, . . . search users *might* see . . . improved privacy protections[.]" Ex. 1 at 18 (emphasis added). He describes DuckDuckGo as a search engine that "promotes itself as offering greater protection for personal data and privacy than Google and Bing" and suggests that "unlike those firms, it does not store user search histories." *Id*. at 26. Professor Baker cites documents which

[1] Professor Fox measured the effect on Google search quality of having query scale reduced. He ran an experiment that "tested two different reduced samples: (1) a sample equivalent to Microsoft's current stream of click-and-query data (Low Mobile) and (2) a sample that used the same amount of desktop training data as in the first sample but doubled the amount of mobile click-and-query training data available (High Mobile)." Ex. 12 at 6. "Using a number of quality metrics, the experiment then compared the search quality resulting for the Low Mobile and High Mobile samples with" results with 100% query volume. *Id.* The DRE showed that "when the data used to train key Google ranking algorithms is reduced to reflect Bing's smaller scale dataset, the decrease in search quality is considerably smaller than Google's current 'lead' over Bing." *Id.*

he asserts show Google offers "less privacy protection" than others. Ex. 3 at 99–100 & n.533. But he has neither expertise in data privacy nor experience or knowledge about the various privacy offerings of any search engine.

ARGUMENT

In deciding a motion to exclude expert testimony, the Court acts as a "gatekeep[er]" to ensure that the requirements of Rule 702 are met. *See Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 597–98 (1993). Plaintiffs bear the burden of establishing Professor Baker's qualifications and the admissibility of each of his opinions. *Sykes v. Napolitano*, 634 F. Supp. 2d 1, 6 (D.D.C. 2009). Under Rule 702, an expert must have "scientific, technical, or other specialized knowledge [that] will help the trier of fact to understand the evidence or to determine a fact in issue[.]" Fed. R. Evid. 702(a). Moreover, the testimony must be "the product of reliable principles and methods," and be "based on sufficient facts or data[.]" Fed. R. Evid. 702(b)–(c).

Plaintiffs fail to meet their burden with respect to Professor Baker's opinions on scale and privacy.

I. Professor Baker's Opinions About Search Engine Scale Should Be Excluded.

Professor Baker's opinions regarding search engine scale fail to meet Rule 702 for multiple reasons. First, Professor Baker is an economist, not an engineer or computer scientist, and he has no relevant experience regarding scale in the highly-technical search engine industry. Second, even if Professor Baker were qualified, his opinions consist mainly of improper narration of documents and testimony from actual industry participants that is not helpful to the fact finder. Third, Professor Baker's rebuttal to Professor Fox's DRE is unreliable and not helpful given his lack of relevant expertise and his reliance on [redacted]

[redacted]. For these reasons, Professor Baker's scale opinions should be excluded.

A. Professor Baker is not qualified to opine on search engine scale.

"The burden is on the proponent of the testimony to show by a preponderance of the evidence that the proffered expert witness is qualified." *Sykes*, 634 F. Supp. 2d at 6 (citing *Meister v. Med. Eng'g Corp.*, 267 F.3d 1123, 1127 n.9 (D.C. Cir. 2001)). Expert testimony needs "a reliable foundation," *Daubert*, 509 U.S. at 597, and must be "'properly grounded, well-reasoned and not speculative before it can be admitted,'" *Estate of Gaither v. District of Columbia*, 831 F. Supp. 2d 56, 62 (D.D.C. 2011) (quoting Fed. R. Evid. 702 advisory committee's note to 2000 amendment). Moreover, if an expert relies "primarily" on his "experience" as the alleged basis for his qualification to testify, the expert's qualifications must be closely related to the subject matter in question. *See id.* at 68–69.

Plaintiffs cannot point to any training, education, or experience as the basis for Professor Baker's opinions on scale in the search engine industry. When asked in his deposition whether he considers himself "an expert in how search queries' scale impacts the quality of search engine results," he admitted, "[t]o the extent that's a field, it would be part of a -- to the extent that's a computer engineering field, I'm not an expert in that." Ex. 4 at 336:20–25. Professor Baker is an economist by training, Ex. 1 at 19, unlike many of the deponents and computer science experts cited above, who have Ph.D.s or other graduate degrees in a relevant field, such as computer science. He therefore has no relevant experience in computer science to support his affirmative opinions, and certainly no experience that would allow him to provide a rebuttal opinion to actual computer science experts like Professors Frieder and Fox.

Take, for instance, his response to Professor Fox's data reduction experiment: Professor Baker asserts that the "volume of clicks and queries matters because [redacted]

[redacted]" Ex. 3 at 62. He explains, "[w]ith few user interactions, it is difficult for a search engine to provide good results, even for the simplest queries and particularly for queries asked occasionally but infrequently (so-called 'body' queries) and queries that are complex and rare (so-called 'tail' queries)." *Id*. at 63. But Professor Baker has no experience that would allow him to assess these highly technical questions, such as whether or by how much the quality of Bing's search algorithm would improve with more user data. Unsurprisingly, then, each of these statements is supported solely by cherry-picked citations to depositions and [redacted]. In other words, whereas Professor Fox uses his expertise to design and run an analysis of the effect of *marginal* scale, Professor Baker seeks to "rebut" this analysis based solely on documents and fact witness statements that he has no expertise to interpret.

While Professor Baker's credentials as an economist may qualify him to serve as an expert on certain economic matters, by his own admission, there is nothing in his educational or professional background that would qualify him to serve as an expert on the issue of how search engine quality and performance is impacted by additional user data. *See Arias v. DynCorp*, 928 F. Supp. 2d 10, 17 (D.D.C. 2013) ("Although Dr. Wolfson has impressive credentials, the plaintiffs have not demonstrated how his academic and professional experiences make him qualified to testify" about one of the issues before the court.). In his deposition, Professor Baker asserted that although he has "never done some sort of -- any computer science like analysis of search query scale and quality," he has experience with "just in general, the concept of diminishing returns to scale." Ex. 4 at 354:5–355:4. That one is an economist does not make one expert in anything that touches commerce. *See United States v. Second Chance Body Armor, Inc*., 289 F. Supp. 3d 145, 177 (D.D.C. 2018) (warning against experts "without any experience working in or studying the

particular industry at issue."). This principle is of particular importance here, where fact witnesses with technical understanding of the industry have testified that the impact of scale is highly dependent on the technology being employed. *Supra* at pp. 3–4; *see also S.E.C. v. Tourre*, 950 F. Supp. 2d 666, 678 (S.D.N.Y. 2013) (rejecting expert that "does appear to have expertise in the general area of structured finance" because "that is so broad a category as to become meaningless when particularized here to synthetic CDOs, a very specific type of security").

Judge Friedman's opinion in *Second Chance Body Armor* is instructive. There, the court excluded a proffered safety expert's opinions in their entirety because the expert failed to explain how his experience in the tire industry, another "safety-related industry," supported his opinions in the ballistics industry. 289 F. Supp. 3d at 176. Given his lack of experience, the expert relied "solely on documents and testimony from this case." *Id.* The court held that the expert could not offer "opinions about standards generally applicable to the 'safety-related industry' without any experience working in or studying the particular industry at issue." *Id.* That principle is all the more applicable here. While Professor Baker may understand how scale economies work in general—such as how, theoretically, the 1000th widget produced might have a lower marginal cost than the first—this does not make him an expert on how query volume (scale) actually impacts search engine results quality and competitiveness in the highly technical search engine industry.

Because Professor Baker lacks the relevant expertise regarding how scale impacts quality in the search engine field, his opinions on this topic should be excluded.

B. Professor Baker's scale opinions consist mainly of improper narration.

Professor Baker's opinions should be excluded for the further reason that they consist mostly of improper narration of evidence in the case. Courts have consistently held that it is "inappropriate for experts to become a vehicle for factual narrative." *See, e.g.*, *Tourre*, 950 F. Supp. 2d at 675. That is because the fact finder "is entirely capable of reviewing the testimony of

[ ] witnesses and drawing conclusions" themself. *United States ex rel. Landis v. Tailwind Sports Corp.*, 2017 WL 5905509, at *15 (D.D.C. Nov. 28, 2017) (citing *United States v. Mitchell*, 49 F.3d 769, 780 (D.C. Cir. 1995)). And "[a]cting simply as a narrator of the facts does not convey opinions based on an expert's knowledge and expertise; nor is such a narration traceable to a reliable methodology." *Tourre*, 950 F. Supp. 2d at 675; *see also Highland Cap. Mgmt., L.P. v. Schneider*, 551 F. Supp. 2d 173, 183 (S.D.N.Y. 2008) (excluding "factual narrative of events" by expert where expert had "no personal knowledge of these facts and they are lay matters that the [fact finder] is capable of understanding and deciding without [expert] testimony"); *In re Rezulin Prods. Liab. Litig.*, 309 F. Supp. 2d 531, 551 (S.D.N.Y. 2004) (expert should not be permitted to provide "narrative of the case" which is "properly presented through percipient witnesses and documentary evidence").

The vast majority of Professor Baker's "opinions" on search engine scale consist of him purporting to interpret documents and testimony from industry participants, without offering any scientific expertise in connection with such narrations. *See* Ex. 1 at 62–64. And in his deposition, Professor Baker admitted that his rebuttal to Professor Fox is "not based on some sort of computer science expertise that I have. It's based on the materials that I referenced," including [redacted]. Ex. 4 at 343:2–24.

All Professor Baker provides is Plaintiffs' gloss on the documents and testimony, which is not helpful to the Court. *See In re Rezulin Prods. Liab. Litig.*, 309 F. Supp. 2d at 551 (rejecting expert narration of facts where "the glosses that [expert] interpolates into his narrative are simple inferences drawn from uncomplicated facts that serve only to buttress plaintiffs' theory of the case"); *Sykes*, 634 F. Supp. 2d at 8 (expert not qualified under Rule 702 where expert did "not offer 'expert' testimony based on his years of experience" but instead "advocate[d] for the Plaintiff

rather than providing expertise to the fact-finder"). The Court should reject such opinions, and instead hear directly from the witnesses themselves.

Professor Baker's narration opinions should therefore be excluded.

C. Professor Baker's rebuttal of Professor Fox's DRE experiment is based on a necessarily incomplete interview and is unreliable.

Professor Baker seeks to offer a critique of Professor Fox's DRE study. Asked in his deposition how he is "qualified to evaluate Dr. Fox's study," Professor Baker said his opinion is "based on my understanding that 'Google's search algorithm has benefited from a long series of improvements over many years, each one of which benefited from Google's scale advantage at the time.'" Ex. 4 at 343:2–12. And he stated that the basis for that understanding was one (1) historical Google document and [redacted]. *Id.* at 343:13–17. As discussed, *supra*, the Court is perfectly able to review that document and evaluate testimony [redacted] without Professor Baker's gloss. His rebuttal of Professor Fox should therefore be excluded on that basis.

But Professor Baker's critique should be excluded for an additional reason: [redacted]

[redacted]. And, therefore, Professor Baker's reliance on [redacted] renders his rebuttal opinion likewise speculative and unreliable. *See Washington v. Kellwood Co.*, 105 F. Supp. 3d 293, 319–23 (S.D.N.Y. 2015) (excluding expert's opinion as unreliable when his sole source for the foundational part of analysis was speculative deposition testimony)

[redacted]

[redacted]

That is, although Professor Baker acknowledges that he does not have the expertise necessary to critique Professor Fox's DRE experiment himself, he asks the Court to nonetheless allow him to provide an expert opinion on the topic based on [redacted]. To state the obvious: that is not "traceable to a reliable methodology." *Tourre*, 950 F. Supp. 2d at 675. Google presumes that [redacted]. But without a connection to the material facts of the study, how could [redacted] be helpful to the Court in assessing Professor Fox's study? If the Plaintiffs wanted to present computer science expertise to rebut Professor Fox's DRE analysis, they should have hired one as an expert—as DOJ plaintiffs did with Professor Oard, *see supra* p. 5. But they cannot bootstrap an expert opinion by having their economist [redacted]

Because Professor Baker's opinion rebutting Professor Fox is unreliable, unhelpful, and not based on his expertise, it should be excluded.

II. Professor Baker's Opinions on Privacy Should Be Excluded.

Plaintiffs cannot meet their burden to show that Professor Baker has any experience or training relevant to assessing the privacy of Google or other search engines. Unlike Professor Tucker, Professor Baker has no expertise in data privacy. Nor does he have any background in computer science. Therefore, he has no specialized experience from which to gauge DuckDuckGo's or Google's offerings in privacy.

Given this lack of experience, Professor Baker's opinions on privacy consist entirely of parroting cherry-picked documents and testimony from the case. *See generally* Ex. 1 at 26; Ex. 3 at 99–100 & n.533. That is not the province of expert testimony. *Tourre*, 950 F.Supp.2d at 675. And it is certainly not the province of an economist who has no expertise in data privacy nor any specialized knowledge about the tradeoffs that search engines must make among privacy and other search-quality factors. [redacted]

The only "analysis" provided by Professor Baker is his unhelpful "gloss" on the documents and testimony that this evidence supports Plaintiffs' claim that Google is less protective of privacy than competitors like DuckDuckGo. *In re Rezulin Prods. Liab. Litig.*, 309 F. Supp. 2d at 551; *see also Sykes*, 634 F. Supp. 2d at 8. But given his lack of experience on privacy matters, Professor Baker is in no better position than the Court to evaluate the evidence on the level of data privacy Google provides to its users (and he certainly is not in a position to rebut actual experts in this area, like Professor Tucker). Professor Baker's opinions thus invade the function of the fact finder. "[I]t is precisely the [fact finder's] function to review the testimony of witnesses and determine what factual conclusions to draw from that testimony." *Tailwind Sports Corp.*, 2017 WL 5905509, at *15.

Because Professor Baker's opinions on privacy are not based on his expertise and are not helpful to the fact finder, they should be excluded.

CONCLUSION

For the foregoing reasons, Google's Motion to Partially Exclude the Opinion of Plaintiffs' Expert Jonathan Baker should be granted, and his scale and privacy opinions should be excluded.

Dated: December 12, 2022

Respectfully submitted,

WILLIAMS & CONNOLLY LLP

By: */s/ John E. Schmidtlein*
John E. Schmidtlein (D.C. Bar No. 441261)
Benjamin M. Greenblum (D.C. Bar No. 979786)
Colette T. Connor (D.C. Bar No. 991533)
680 Maine Avenue, SW
Washington, DC 20024
Tel: 202-434-5000
jschmidtlein@wc.com
bgreenblum@wc.com
cconnor@wc.com

WILSON SONSINI GOODRICH & ROSATI P.C.
Susan A. Creighton (D.C. Bar No. 978486)
Franklin M. Rubinstein (D.C. Bar No. 476674)
Wendy Huang Waszmer (D.C. Bar No. 1631078)
1700 K Street, NW
Washington, DC 20006
Tel: 202-973-8800
screighton@wsgr.com
frubinstein@wsgr.com
wwaszmer@wsgr.com

ROPES & GRAY LLP
Mark S. Popofsky (D.C. Bar No. 454213)
2099 Pennsylvania Avenue, NW
Washington, DC 20006
Tel: 202-508-4624
Mark.Popofsky@ropesgray.com

Matthew McGinnis (admitted *pro hac vice*)
Prudential Tower
800 Boylston Street
Boston, MA 02199
Tel: 617-951-7703
Matthew.McGinnis@ropesgray.com

Counsel for Defendant Google LLC