In Google we trust?

Roberto Burguet[*], Ramon Caminal[*], Matthew Ellman[*†]

October 2013

Abstract

This paper develops a unified and micro-founded model of search and display advertising where consumers search for offline products and for online content published on platforms that display ads. We identify novel incentives for a monopoly search engine to distort its organic and its sponsored search results. First, distorting organic results to make display advertising less effective raises merchants' willingness to pay for sponsored search. Second, underweighting consumer relevance in sponsored search auctions cashes in on higher margin merchants. The interplay between these two incentives and the need to attract search participants determines search bias and welfare. We also characterize how the welfare consequences of search engine integration into ad intermediation or publishing depend on publisher asymmetries and on targeting and market concentration in display advertising.

Keywords: Search engine bias, internet economics, vertical integration, two-sided markets, antitrust. *JEL Classification*s: L13, L41, L82, L86.

[*]Institut d'Anàlisi Econòmica CSIC, and Barcelona GSE.
Emails: roberto.burguet@iae.csic.es, ramon.caminal@iae.csic.es, matthew.ellman@iae.csic.es.

[†]We thank Konstantina Bania, Michael Baye, Stefan Bechtold, Jan Boone, Emilio Calvano, Lisa George, Shane Greenstein, Andrei Hagiu, Sjaak Hurkens, Michael Katz, Yossi Spiegel, Catherine Tucker, Alexander White, Joel Waldfogel and seminar participants at several conferences (Évora ICT, 2013, Tilburg, TILEC III, 2013, Chicago Searle Internet Conference, 2013, Paris SAET, 2013 for useful comments. We also acknowledge financial support from the Spanish Ministry of Science and Innovation (Grant: ECO2011-29663).

"*[W]e expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers.*" - Brin and Page (1998)

1 Introduction

Search engines play a critical gatekeeper role in determining people's access to online information as well as to offline products; indeed they dominate the growing market in web-based advertising. With market shares exceeding 90% in most European countries and a global average above 80%, Google arguably dominates online search in most of the world (State-of-search, 2012). This may reflect widespread trust in Google's motives and ability to deliver reliable search results, but a growing chorus of voices cautions against relying on Google's "do no evil" promise to deliver unbiased results. These critics point to evidence of specific search biases that raise the ranking of Google's own content and services, that is, webpages where Google acts as a "publisher" (see e.g., Edelman and Lockwood, 2011, Edelman and Lai, 2013, Tarantino, 2013, and www.FairSearch.org, or van Loon, 2012, for a legal discussion). Search biases between websites not owned by Google are harder to detect, but a unified model of Google's incentives can indicate where to look for bias and allows to evaluate the welfare implications of Google's expansion into display advertising, either as an ad intermediary (AdSense and AdWords) or as a publisher (e.g., Google Finance, YouTube and Zagat).

In this paper, we develop a micro-founded economic model that integrates the sponsored and organic search literatures. Our analysis features consumers seeking content and products, merchants selling offline products, publishers offering online content (or services), ad intermediaries helping merchants advertise on publisher websites, and a monopoly[1] search engine, that directs consumers to merchants and publishers via its search rankings and we denote G.[2] We model both organic search (where rankings are not paid for) and sponsored search (where a position auction determines the ranking of links) and we characterize G's incentives to distort each type of search result. We derive two types of bias, one affecting organic search and one affecting sponsored search, which operate independently but also interact.[3] We then investigate the bias and welfare consequences of G's expansions into display advertising, as publisher or ad intermediary. Such integrations have both vertical and horizontal features: G's organic search

[1] Alternative search engines may be *"just one click away"* as is often argued, but indirect network externalities, combined with small switching costs, habit effects and delays in detecting reduced search quality, interfere with competitive forces; Argenton and Prüfer (2012) and Etro (2012) model how a search engine with an initial advantage can monopolize the market by learning from past searcher experiences. Brin and Page (1998) also anticipated the quality detection problem: *"Since it is very difficult even for experts to evaluate search engines, search engine bias is particularly insidious."*

[2] The label, "G," fits the motivating application to Google, but the analysis can apply more generally.

[3] Throughout the paper, we define bias relative to what consumers are searching for, though we take account of producer surplus in our welfare analyses.

provides the audience for publishers and their display ads; at the same time, display ads are imperfect substitutes for G's sponsored search ads.[4]

The quality of search affects welfare by (a) matching consumers with products they may wish to buy and (b) matching consumers with content they may wish to consume online, which (c) determines the effectiveness of display advertising and (d) influences surplus appropriation and the investment incentives of the five different groups of actors.

We distinguish between products and content because display advertising is readily tied to online content but not to offline products: publishers can display third-party ads to visitors while they consume content on the publisher's website. So publishers operate in a two-sided market. On the advertising side, their display ads compete with G's (sponsored) search ads, while on the consumer side, G's search engine, as upstream gatekeeper, directs traffic to publishers. It is this twin interaction that gives G both the incentive and the power to interfere with display advertising by distorting organic search as we explain shortly.

In parallel, we distinguish consumer searches by their objective: "content searches" for online goods and "product searches" for offline goods.[5] In principle, consumers could reach both publishers' content and merchants' products via either sponsored or organic search results, but in the equilibrium of our model, consumers only use organic results when seeking content and only use sponsored results when seeking products. This simplifying split reflects two ideas.[6] First, for product searches, G has clear incentives to distort its organic results enough to induce consumers to only click on the sponsored results, because merchants must then sponsor links to get traffic from G.[7] Second, as a rule, publishers do not buy sponsored links, so consumers only use organic results when conducting content searches.[8] This split captures, in a stylized fashion, the findings of Greenspan (2004) and Jansen (2007) that people use sponsored links more than organic results when conducting product searches or "e-commerce search queries", but place more trust in organic results when seeking content (see Hotchkiss et al., 2005, and Jansen and Resnick, 2006). Thanks to this split, we only need to keep track of two types of bias: bias in organic results for content searches and bias in sponsored results for product searches.

[4]We characterize this substitutability of display and search advertising which lies at the heart of recent regulatory controversy; see FTC statement 071-0170 and the contrasting view in the EU report, M.4731.

[5]In Broder's (2002) taxonomy, these correspond, respectively, to "informational" and "transactional" searches which Jansen et al. (2008) estimate to represent 80% and 10% of all searches; the remaining 10% are in a third category, "navigational" searches, where the searcher seeks to find a specific website.

[6]It makes the model highly tractable. We readily account for a less extreme split in section 8.

[7]Brin and Page (1998) allude to this type of bias in the above quotation, as is clear in their airline example. They also note that "*a search engine could add a small factor to search results from friendly companies, and subtract a factor from results from competitors. This type of bias is very difficult to detect but could still have a significant effect on the market.*"

[8]Section 8 discusses explanations for this stylized fact in terms of the transaction costs that publishers with dynamic content would face in bidding, the informational nature of online content and G's ability to use organic results to indirectly commit against distorting search to favor sponsors.

The first potential source of bias for these organic results (in content searches) derives from the substitutability between display and search as advertising channels for merchants. Display advertising therefore poses a competitive threat to G's search ad revenue. Display is a stronger substitute for search advertising when consumers are more attentive, receptive and responsive to ads. This effectiveness of display advertising depends on organic search distortions: (a) the average effectiveness of display ads falls when G distorts organic search towards publishers that are less effective for display advertising, or publishers that, like the BBC and Wikipedia, choose to have no display ads; (b) even with symmetric publishers, distorting organic search worsens the matching between consumers and publishers, and this undermines display advertising because consumers are less attentive to less relevant publishers.[9] When G integrates into publishing or ad intermediation, a second potential source of organic bias derives from the fact that G only internalizes the ad revenues of its own publishers and affiliates – those publishers that pay G for ad intermediation.[10]

The source of incentives to bias sponsored search lies in a potential conflict of interest between consumers and merchants over ranking sponsored links: the merchants most willing to pay for a top position may not be the best option for consumers since net margins and net consumer values need not be perfectly aligned. G's scoring auction for sponsored links can discount the bids of merchants whose products provide lower net consumer value, but G internalizes a share of merchant profits so it may be tempted to underweight consumer value so that less relevant merchants with higher margins can win.[11]

These temptations to bias organic search and sponsored search are naturally tempered by G's need to attract consumers to search via G. In the resulting tradeoff, a profit-maximizing search engine may distort search results relative to the allocation rules that maximize consumer surplus. The incentives to distort organic and sponsored search interact as imperfect substitutes for G: more bias in one type of search lowers the incentive to distort the other. The overall tradeoff depends on parameters such as display ad targeting which raises the substitutability of search and display advertising and exacerbates the incentives for traffic distortion. Clear product search queries always permit perfect targeting of search ads, but ongoing technological innovation is rapidly improving display ad targeting and thereby increasing the importance of the issues we discuss here.

Integration into ad intermediation changes G's distortion incentives. We present three main effects:

[9]This is in line with Google AdSense's advice to publishers that attracting relevant consumers, as well as publishing interesting content, will raise their display ad revenues. Intuitively, consumers are less exposed to ads on on less relevant webs or web domains, since they spend less time there. A growing body of evidence on search using eye-tracking technology also support this idea; see Lorigo et al. (2008) and in particular, Wang and Day (2007). Ellman and Germano (2009) and Wilbur (2008) offer further evidence and richer views on the relationship between content and advertising effectiveness.

[10]Own-content bias is, with affiliate bias, the simplest and most recognized search bias in regulatory circles.

[11]So the model only partially supports Google's claims about its sponsored search position auctions - see Google's chief economist, Hal Varian's YouTube video or Varian (2007) for a description of how the scoring auction should weight each merchant's bid by merchant quality and relevance to the searching consumer.

(1) integration with monopolization of the ad intermediation market improves the reliability of G's search results, because G then internalizes profits from display advertising; (2) on the other hand, integration without full monopolization biases G's organic search to favor publishers that deal with G as ad intermediary. In particular, we identify conditions under which non-integration generates higher total welfare and higher consumer surplus than does integration with partial monopolization. In addition, allowing for asymmetry, we show: (3) even with full monopolization, integration can have negative consequences for organic search and for total surplus when publishers vary in effectiveness as platforms for display advertising.[12]

Advertising revenues are fundamental to the business models of most web-based publishers, including the much-beleaguered news media. We show that G's integration into ad intermediation may sharply reduce the share of advertising surplus that publishers can appropriate. This presents a serious concern: in an extended model with investments, this would discourage publishers from investing in quality and new content.

Internet-based trade and search engine incentives are active research fields in economics.[13] Our unified model covers both organic and sponsored search as well as explicit competition between platforms offering advertising opportunities to merchants. Early work in the literature studied sponsored search alone.[14] Athey and Ellison (2011) and Chen and He (2011) are important precursors to our study. They show how position auctions induce positive self-selection when merchants have private information about their relevance to consumers. In these papers, more relevant merchants generate a higher probability that consumers click on their links (in a given position) through to the merchant – this is the "click-through rate" or CTR. Hagiu and Jullien (2011) derive a similar self-selection effect in a model (abstracting from search auctions) where merchants differ in the probability that a click leads to a purchase – this is often called the "conversion rate" or CR. In both the CTR and CR dimensions, merchant and consumer preferences are aligned, because consumers seek merchants that will attract them to click-through and buy. We instead analyze a setting where merchant and consumer interests conflict. This generates a straightforward conflict between the search engine and consumers.

By contrast, prior work has relied on asymmetric information to explain why search engines might bias product search. First, in papers such as Athey and Ellison (2011), the search engine may distort

[12] We develop similar results for publisher integration; direct control over display advertising can lead to minor variations.

[13] Baye and Morgan (2001), a milestone in this literature, examines information gatekeepers on the web, modeling price-comparison sites but is relevant to search engines; the gatekeeper's two-sided pricing strategy affects consumer search and thereby affects merchant pricing in a downstream homogeneous product market.

[14] On equilibrium bidding by merchants in position auctions, see e.g., Lehaie (2006), Edelman and Ostrovsky (2007), Edelman and Schwarz (2010), Edelman, Ostrovsky and Schwarz (2007), Varian (2007), Borgers et al. (2007). On auction design, see e.g., Liu, Chen and Whinston (2010) and Athey and Ellison (2011) where scoring auctions with reserve values are designed to extract merchant rents. The general literature on merchants buying prominence is also relevant as top search auction positions afford prominence; see e.g., Armstrong, Vickers and Zhou (2009) and Armstrong and Zhou (2011).

the ranking or inclusion of merchants in its effort to extract rent from privately informed merchants. Second, in Athey and Ellison (2011) and Hagiu and Jullien (2011), consumers have private information about their trade values from different merchants and the search engine may distort its results to increase consumer search which has a positive externality on merchants and the search engine. Finally, a third strand of the literature shows how search distortions can affect the determination of merchant values via product market price competition (see e.g., Chen and He, 2011, Eliaz and Spiegler, 2011, Hagiu and Jullien, 2011, White, 2013, Xu, Chen, and Whinston, 2010 and 2011). Our model of conflicting interests in sponsored search is simpler, being a direct consequence of imperfect alignment between merchants' margins and consumers' net benefits.

A number of papers have looked into organic search bias and the effect of integration. Hahn and Singer (2008) provide an informal law and economics analysis of the Google-DoubleClick merger of 2007, also treated in White (2013). White develops the key insight that organic results can interfere with sponsored search; in his model, search distortions serve to reduce merchant competition in a downstream market, thereby raising sponsored search revenues, and White discusses how integration may reduce this bias. On the other hand, White's paper lacks a micro-founded model of search and advertising.[15] More recently, in independent work, de Cornière and Taylor (2013) also examine organic search bias and integration. Their reduced-form model assumes strategic substitutability between product advertising on G's and publishers' websites, which induces a motive for bias in (organic) search. Under this assumption, they look at the impact of integration with one of two asymmetric publishers.[16] Our micro-founded framework explicitly models both the search in search advertising and the platform competition between G and publishers for merchants. In particular, endogenizing merchant demand allows us to derive (in terms of parameters such as display ad targeting) the degree of substitutability betweeen these advertising channels, a critical factor in ongoing regulatory debates and in determining organic bias. In addition, we derive a bias in sponsored search, which interacts non-trivially with organic search bias. For instance, an improvement in publishers' targeting technologies can reduce the quality of organic results, which is quite intuitive, but may alternatively worsen the sponsored results, which is much less intuitive; for another example, an increase in consumer value of accurate content results can increase sponsored bias as well as decrease organic bias.[17]

[15] Concretely, White's (2013) merchants sell identical products and cannot choose between display and search advertising: "right-side" merchants can only buy search ads, while "left-side merchants" get free advertising. Our organic bias results from competition between these ad platforms for (heterogenous) merchants choosing display and/or search advertising.

[16] White and Jain (2012) also study websites, such as a search engine and a publisher, that earn revenues from advertising in a reduced-form model, but they do not model consumer search, instead endogenizing ad levels with nuisance effects.

[17] Tarantino (2013) adds a further twist by examining how a general search engine integrated with a vertical search engine, such as a web-based travel intermediary, distorts traffic towards this owned intermediary. Notice that a travel intermediary typically sells travel products to consumers and sponsors travel-related product searches on general search engines, just as do the merchants in our model.

The paper is organized as follows. Section 2 presents the baseline model and characterizes the social optimum, before analyzing the equilibrium of the game with full separation in section 3, with monopolizing integration into ad intermediation in section 4, and partial monopolization in section 5. In section 6, we allow for publisher asymmetries in ad effectiveness and section 7 treats integration with publishers. In all cases, we analyze welfare and surplus implications. Section 8 discusses assumptions and extensions and we conclude in section 9, gathering proofs in the Appendix.

2 The baseline model

Here we present the essential elements of the baseline model, postponing to Section 8, the motivation and endogenization of our main assumptions, and some possible generalizations. There are five types of agent: consumers, merchants, publishers, intermediaries (between publishers and merchants) in the display advertising market, and a search engine, G.

Products, content, and consumer demand. A mass one of consumers, indexed by i, value specific varieties of offline and online goods. Each consumer has unit demand for offline goods and also unit demand for online goods. Offline goods, called "products," vary by category $j \in \{1, 2, ..., J\}$ and type $k \in \{1, 2\}$, giving $2J$ products defined by pairs (j, k). Consumer i only values one category, denoted $j(i)$, and one unit of product $(j(i), k)$ gives i a net benefit v_k where $0 < v_2 < v_1$; so $(j(i), 1)$ and $(j(i), 2)$ are i's "best-match" and "second-best" products. All other products or additional units imply a net loss.

Online goods, called "content" are available on publishers' websites. Each of N publishers has exactly one website with unique content, so publishers, content and websites share an index, n. Each consumer i has a favorite or "best-match" content, $n(i)$, that generates net utility, $u > 0$, while any other content, $n \neq n(i)$, generates zero net utility and further units imply a net loss. We treat the symmetric case where each product category and each publisher's content interests the same fraction of consumers, $\frac{1}{J}$ and $\frac{1}{N}$, respectively.

Merchants and publishers. Merchants each sell one product, which they make available on their websites. We call a merchant type k if its product is type k; such merchants earn a unit margin m_k, where we assume $m_1 < m_2$. This homes in on the possible conflict of interest between consumers and merchants that arises whenever consumer values and merchant margins are imperfectly aligned: $m_1 < m_2$, whereas $v_1 > v_2$. To simplify the welfare analysis, we assume $m_1 + v_1 > m_2 + v_2$, so the social optimum has only type 1 transactions. Two merchants produce each product, implying $4J$ merchants and ensuring competition for all advertising opportunities.[18] Publishers make their content available

[18]Having at least two merchants per product implies zero merchant profits in equilibrium, which greatly simplifies the

free of charge to consumers, but exclusive to their own websites where they can expose the consumers to display advertising.

The search engine. Consumers do not know the identity of the publishers and merchants offering their preferred online and offline goods. They can use G to search for content and then for products.[19] To conduct each type of search, they type in a query, consisting of a set of keywords. G can perfectly interpret these queries and responds by providing a set of results in the form of links to publishers' or merchants' websites. These links may be paid for by the party receiving the link ("sponsored results") or not paid-for ("organic results"). As motivated in the introduction, we assume G provides only organic results for content searches and only sponsored results for product searches. Moreover, G provides just one link for each type of search; see section 8 on endogenizing these assumptions. So we can characterize G's response to content queries by the probability, denoted r^O, that G's link leads to the consumer's best-match publisher, $n(i)$, where G chooses $r^O \in [0,1]$ and otherwise links to some other publisher. We characterize how G responds to product queries under (sponsored) search advertising just below.

Advertising. Merchants have two channels for advertising. They can sponsor search links on G's search results page to promote their products to searchers; we call this "search advertising." Merchants can also advertise by buying display space on publishers's sites; we call this "display advertising." So G's search results page and publishers' websites represent competing platforms or channels for advertisers to reach consumers. G and publishers sell their respective ads, search and display, through variants on a second-price auction where bidding determines the "pay per click" price, denoted "PPC". G sells its search ads directly to merchants. Publishers typically sell display ads to merchants with the help of ad intermediaries that serve merchants' ads onto the display areas of publishers' websites; these intermediaries may also organize publishers' display ad auctions, as in an ad exchange, but we describe the equivalent case where publishers run their own auctions.

G offers one search ad slot - a single sponsored link to a merchant's website - every time a consumer i enters a product query. Consumers always click on the search ad link during a product search, because there is only one link and it might be useful. So this "click-through" rate to the merchant's website is one. Similarly, each publisher n offers one display ad slot for a link to a merchant's website, whenever a consumer i visits n's website to consume its content, but these links have a lower click-through rate, because consumers on publisher websites, being focused on consuming content, may not notice and click on the link: the rate is $\alpha \in (0,1)$ if i is visiting her best-match publisher $n(i)$ and only $\alpha\beta$ if i is on any other publisher's web. The reduction factor $\beta \in (0,1)$ captures the idea that consumers spend less

analysis, but this competition is not a critical assumption.

[19] This homes in on the generic feature that consumers engage in some content search before satisfying all possible product demands; see section 8.

time on webs with less relevant content, reducing the probability that they notice and click on the ad; see footnote 9 and section 8.[20]

Search ads also permit better targeting than display ads, because G observes and can reveal each product searcher's query to the merchants who can then infer the consumer's category interest, $j(i)$.[21] By contrast, publishers, aided by ad intermediary targeting technologies, only observe a signal $s(i) \in \{1, 2, ..., J\}$ of each visiting consumer's product category interest. This signal correctly indicates i's preferred product category with probability σ: formally, $s(i) = j(i)$, with probability σ and otherwise points to a random distinct category.[22] If publishers (or their ad intermediaries) reveal these consumer signals to merchants, merchants can target their display ads with precision σ.

Display advertising auctions. For each consumer visit, publishers conduct second-price auctions of the PPC for displaying an ad, contingent on the signal $s(i)$ of the consumer's preferred product category.[23] Merchants submit bids and the publisher displays one highest bidding merchant's ad to this consumer. Publishers have no costs except possible charges for ad intermediation. These charges are zero in the case without integration, because intermediaries have zero costs and we assume free entry. So we need only model ad intermediaries explicitly when the search engine owns one.

Search advertising. For each consumer who conducts a product search, G uses a weighted second-price auction to allocate the single sponsored result that G shows the consumer when she types in her product query.[24] Merchants observe the query and, anticipating that, in equilibrium, each pair of identical merchants make identical bids; we denote by b_k the bids of the relevant type k merchants - namely, those selling $(j(i), k)$. Merchants selling products irrelevant to a query never bid a positive amount, so G has no need to discount bids based on category relevance; consumer-merchant interests are aligned in this dimension. However, the scoring auction does discount merchants based on type, which corresponds to the "quality" described in Varian (2007) since $v_1 > v_2$. Concretely, G discounts type 2 merchant bids b_2 by a factor $\mu < 1$ which G chooses. The winner of the auction is determined by comparing the weighted bids, b_1 and μb_2. The PPC rate is set equal to the lowest bid that would have allowed the winner to win the auction. In equilibrium, the auction is tied, at least between identical

[20] We rule out that G simply hides all content sites by giving no result at all. This would seem to block all display advertising and might be attractive to G, but full hiding is unrealistic and readily excluded by the arguments against blatant search distortions in section 8. Also participation concerns lead G to prefer to give some result if having no result has sufficiently negative net utility consequences for searchers or if searchers can access (negative net utility) back-up content sites with higher ad-effectiveness.

[21] All consumers prefer type 1 merchants so this aspect of consumer demand is anyway common knowledge.

[22] Targeting precision satisfies $\frac{1}{J} \leq \sigma \leq 1$ since $\sigma = \frac{1}{J}$ if the ad intermediary has no information on consumer preferences and $\sigma = 1$ if the ad intermediary can identify the consumer's best-match category with probability one. $n(i)$ and $s(i)$ are independent.

[23] Rather than reveal their signals to merchants in real-time, publishers and G in its search auctions, make their information available for interaction with merchants' bidding strategies in automated auctions. But it is equivalent to describe as if merchants bid after observing the information.

[24] This is variously called a position, paid-for placement, sponsored keyword or search auction.

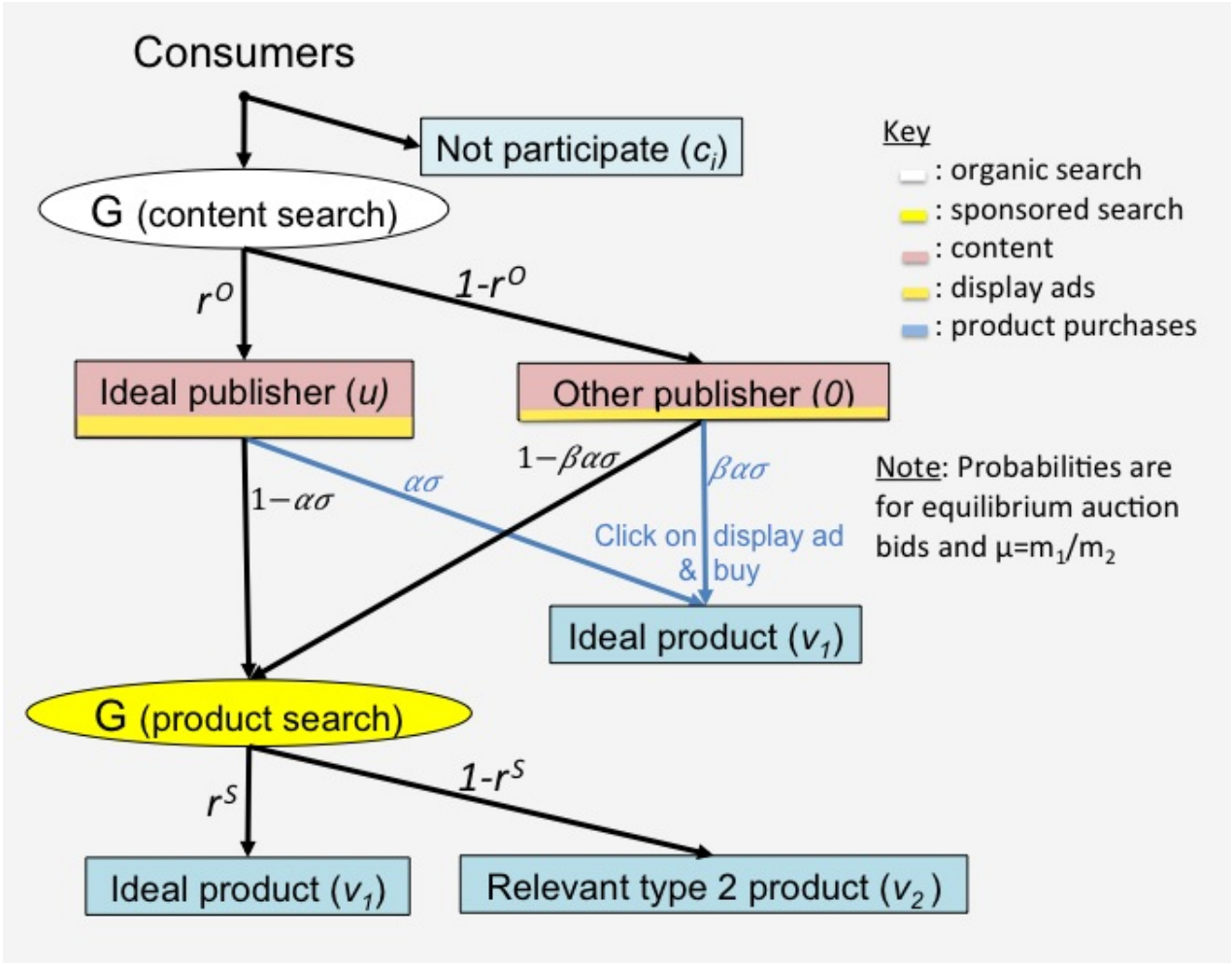
merchants, so the winner pays her bid. If $b_1 > \mu b_2$, a randomly chosen type 1 merchant wins and the PPC rate is b_1. Conversely, if $b_1 < \mu b_2$, a type 2 merchant wins, paying a PPC of b_2. If the two types tie, implying four possible winners, G applies a tie-breaking rule that favors a type 1 merchant with probability r^S and a type 2 merchant with probability $1-r^S$. In equilibrium, we will see that G chooses μ so that the types always tie; r^S then represents the "reliability" of sponsored search.

Consumer participation. We model consumers' participation in using G's search engine as a single decision that depends on the overall reliability of sponsored and organic results.[25] Each consumer i has a joint cost c_i of using the search engine for both content and product queries; this includes the foregone expected benefits from alternative search. We assume that c_i is an independent draw from a continuous random variable on $[0, c_H]$ with density function $f(c)$ and cumulative distribution function $F(c)$, such that the reverse hazard rate, $H(c) = \frac{f(c)}{F(c)}$, is decreasing.[26] Each consumer observes her cost c_i privately and prior to deciding participation. Gross of her direct plus opportunity costs given by c_i, i's expected gain from participating in online search is the sum of expected gains from consuming online content ($r^O u$) plus offline products, found via a display ad during content search or via a search ad during product search. The highest possible such gain is $u + v_1$. We assume $c_H > u + v_1$ so that consumer participation in online search is interior in any equilibrium.

The timing. In the first stage, G and publishers announce their auction rules, including design variables, μ, r^S and r^O. In the second stage, merchants choose their bidding strategies for both search and display advertising auctions. In the third stage, consumers decide whether or not to use the search engine. If they do participate in search on G, they type in their query for content and can visit one publisher's website. While consuming the publisher's online content, they may be attracted to click on its displayed ad through to a merchant's web where they may buy the merchant's product. Then they either leave the market or they type in a product query and can visit one merchant's web where they can buy that merchant's product. Merchants, publishers, ad intermediaries, and consumers always observe the outcomes of previous stages. Figure 1 depicts the timing and indicates the probabilities for consumer traffic and trade alternatives that arise when $\mu = \frac{m_1}{m_2}$ and merchants make equilibrium bids. Publisher ad space and G's sponsored search results are marked in yellow to indicate their horizontal relationship. The vertical relation between G and publishers is also visible in the arrows from content search on G to publishers: G's organic results, marked in white, play the role of "information gatekeeper," directing traffic to publishers.

[25] Section 8 rationalizes this in terms of limited adaptation of search technique to specific search objectives, owing to simple habits or heuristics, or coarse learning about search reliability.

[26] This ensures that G faces an increasing 'marginal factor cost' of attracting the consumer base it "sells" on to merchants.



The social planner's problem. For a benchmark, we consider a social planner who maximizes the sum of all agents' surpluses by controlling how the search engine matches consumers with merchants and publishers, how publishers and intermediaries allocate display ads among merchants, and which consumers participate in online search. For any given participation level, the best possible outcome from this total surplus perspective is for each consumer to consume her best-match content and one unit of her best-match product; recall $m_1 + v_1 > m_2 + v_2$. This is feasible: G can send each consumer to the best-match publisher, $n(i)$, and can send any product-searching consumer to a best-match merchant, that is, one producing $(j(i), 1)$. This is also necessary for efficiency. First, content search must be undistorted since consumers have no alternative way to find content. Second, product search must be undistorted, because the alternative channel, display advertising, at best permits consumers to find their best-match products with probability $\alpha < 1$.[27] Abusing notation in anticipation of the equilibria where merchant types tie, we momentarily let r^O denote the probability that G sends content searchers to

[27] In the baseline model, display ads are redundant because product search has no imperfections, nor added costs given participation; see sections 6 and 7 for an extension. Display ads are also harmless because consumers ignore irrelevant or type 2 display ads, but see 8 on compulsive consumers.

their best-match publishers, while r^S is the probability of sending product searchers to their best-match merchants – a type 1, relevant merchant. So we have,

Proposition 1 *Total surplus maximization requires the search engine to allocate traffic with no distortion from the consumer's ideal; in the first-best,* $r^O = r^S = 1$.

This proposition also holds in the constrained scenario where the planner cannot control consumer behavior (search participation and trade). In fact, the need to attract these consumers, who neglect the positive externality of their participation on producer surplus, only reinforces the planner's incentives to not distort search.

3 Equilibrium analysis under vertical separation

Throughout the paper, we solve for subgame perfect equilibria in undominated strategies. Consumers make many decisions but most are immediate once stated. As already noted, consumers always click on the single link after entering either type of query; that is, they follow G's "recommendation." Similarly, consumers only ever buy a product from their category of interest and they never buy more than one unit overall. In the product search stage, consumers have no subsequent chance to find attractive products, so they buy the advertised product whether type 1 or type 2, provided it is from the relevant category, which it is in all equilibria of the baseline model. Anticipating this, in the prior content search stage, consumers only buy a displayed product if it is type 1 as well as relevant.[28] Consumers omit the product search if a display ad satisfies their demand for products since they only demand one unit, but they always gain from content search, if they paid their cost of participating in online search. Participation is the only remaining non-trivial consumer decision and is characterized below.

Search auctions. We begin by studying the optimal design of sponsored search auctions and merchants' equilibrium bidding strategies. Consumers who did not purchase offline goods through display advertising enter their product query in the search engine. Four merchants, two of each type k, are potentially relevant and correctly anticipate that every click on their search ad leads to a purchase, so type k's are willing to pay a PPC of m_k to appear in the single slot of G's sponsored search results; merchants' bids have no impact on their alternative sales options. As in unweighted second-price auctions, each merchant's unique weakly dominant strategy is to bid her willingness to pay, $b_k = m_k$:

Lemma 1 *For any* μ, r^O, r^S, *the strategy profile* $(b_1, b_2) = (m_1, m_2)$ *is the unique equilibrium.*

[28] A type 2 display ad is less attractive than continuing to a product search which offers some chance of a relevant type 1 product; $r^S v_1 + \left(1 - r^S\right) v_2 \geq v_2$. If $r^S = 0$, consumers would also be willing to buy type 2 products via display ads, but this cannot occur in equilibrium; the expectation of type 2 display purchases would lead G to set r^S slightly above 0 in order to induce type 1 display purchases, thereby raising consumer participation.

So if $\mu < \frac{m_1}{m_2}$, type 2 bids are so discounted that a type 1 merchant always wins and traffic is efficiently allocated. Conversely, if $\mu > \frac{m_1}{m_2}$, type 2 merchants always win and product search is inefficient. Discounting type 2 merchants by the precise weight, $\mu = \frac{m_1}{m_2}$, equates the two types of merchant's effective willingness to pay, allowing G to use its tie-breaking rule r^S to fine-tune the probability, then equal to r^S, that a type 1 merchant wins the position auction. G need only consider this last case, since G can always set $r^S = 1$ and 0 to generate outcomes equivalent to higher and lower μ, respectively; section 8 discusses alternatives to this auction. Competition among merchants implies that, in equilibrium, the winner always pays its willingness to pay. Summarizing:

Lemma 2 *G optimally sets auction weight $\mu = \frac{m_1}{m_2}$, so that a relevant type 1 merchant wins the auction, paying m_1 per click, with G's chosen probability r^S, and a relevant type 2 merchant wins, paying m_2 per click, with probability $1 - r^S$; G's average revenue equals the average margin, denoted $M\left(r^S\right)$, on sponsored-search-mediated sales: $M\left(r^S\right) = r^S m_1 + \left(1 - r^S\right) m_2$.*

Display auctions. We now turn attention to the second-price auctions for display advertising that take place whenever a participating consumer lands on a publisher's webpage, during her first stage of search, her content search. Merchants compete by bidding their willingness to pay per click. They anticipate zero rents from any product searches, so merchants have no opportunity cost of winning a display ad, nor any indirect benefit. Clearly, type 2 merchants cannot gain from bidding for display ads, given that consumers only ever buy type 1 products in the content stage. The merchants indicated by the targeting signal – those selling $(s(i), 1)$ – anticipate that if they win a display ad, a fraction σ of clicks will yield sales, so they bid σm_1, while other type 1 merchants only bid $\frac{1-\sigma}{J-1} m_1$. The merchants selling $(s(i), 1)$ always win, but beyond the special case of perfect targeting, their sales per click or conversion rate, σ, is still strictly below the full unit rate in search advertising. Publishers have no incentives to use a weighted auction to allocate their display ad slots, since type 2 merchants never bid.[29] So we have,

Lemma 3 *Publishers optimally conduct unweighted auctions and the type 1 merchants indicated as most relevant by the targeting technology bid σm_1 per click and one of them wins.*

Recall that a consumer visiting a publisher website is attracted by a display ad with probability α if visiting her favorite publisher and $\alpha\beta$ if visiting any other publisher. So in equilibrium, the average fraction of clicks on display ads is $\alpha e\left(r^O\right)$, where $e\left(r^O\right) = r^O + \left(1 - r^O\right)\beta$ and is increasing in r^O since

[29] In any case, publishers would internalize little of the consumer participation benefit from type 1 display ads since N is usually large.

$\beta < 1$. A fraction σ of these clicks end up in a transaction, so the fraction, η, of participating consumers who buy via display advertising (always their best-match product) is given by,

$$\eta = \sigma\alpha e\left(r^O\right) = \sigma\alpha\left(r^O + \left(1 - r^O\right)\beta\right).$$

Note that η increases with the quality of the targeting technology, σ, and with the reliability of organic results, r^O. Letting X denote consumer participation, the mass of consumers performing a product search is $X\left(1-\eta\right)$.

Consumer participation and continuation equilibrium. Consumer participation is determined by the expected benefit from search participation, which we denote by $\bar{c}$; all consumers with a lower cost c_i will participate. So $X = F\left(\bar{c}\right)$, where $\bar{c}$ is the sum of three expected net utilities from i consuming, respectively, online content (net value u or 0), her best-match product (net value v_1) and her second-best product (net value v_2):

$$\bar{c} = r^O u + v_1\left[\eta + \left(1-\eta\right)r^S\right] + v_2\left(1-\eta\right)\left(1-r^S\right). \tag{1}$$

The probability in brackets of finding her best-match product sums the probabilities η via display advertising and $\left(1-\eta\right)r^S$ via search advertising, whereas she only ever consumes her second-best product via search advertising, in the probability $\left(1-\eta\right)\left(1-r^S\right)$ event that she neglects display ads during content search and G shows a type 2 merchant. Note that $\bar{c}$ increases with r^O, r^S and σ. So if G distorts traffic, by setting low values of r^O and/or r^S, it pays the cost of reduced consumer participation. The two instruments, r^O and r^S, play a similar role in encouraging participation and a high value of one reduces the sensitivity of participation to the other; that is $\frac{\partial^2\bar{c}}{\partial r^S \partial r^O} < 0$. Drawing all this together, we have:

Proposition 2 *In the unique continuation equilibrium following any* $\left(r^O, r^S\right)$*, merchants make zero profits, each publisher earns* $\Pi_n = \frac{F(\bar{c})}{N}\eta m_1$*, and the search engine earns* $\Pi^G = F\left(\bar{c}\right)\left(1-\eta\right)M\left(r^S\right)$.

Traffic management. Of the three positive factors constituting G's profit, r^O increases the first (participation, $F\left(\bar{c}\right)$) and decreases the second (the fraction who product search, $1-\eta$), while r^S increases the first and decreases the third (the average margin of search-based trades, $M\left(r^S\right)$). So there is a simple trade-off for organic search: lowering r^O shifts trades from display to search platform, raising $1-\eta$, but increasing r^O attracts search participation, $F\left(\bar{c}\right)$. Similarly, there is a simple trade-off between raising r^S to attract participation, and lowering r^S to raise the average margin of search-based trades. Mathematically, the first-order conditions for maximizing G's profits with respect to r^O and r^S are:

$$\frac{\partial\Pi^G}{\partial r^O}\frac{1}{M\left(r^S\right)} = f\left(\bar{c}\right)\frac{\partial\bar{c}}{\partial r^O}\left(1-\eta\right) - F\left(\bar{c}\right)\frac{d\eta}{dr^O} = 0, \tag{2}$$

$$\frac{\partial \Pi^G}{\partial r^S}\frac{1}{1-\eta} = f\left(\overline{c}\right)\frac{\partial \overline{c}}{\partial r^S}M\left(r^S\right) + F\left(\overline{c}\right)\frac{dM\left(r^S\right)}{dr^S} = 0. \tag{3}$$

which can be rewritten using the reverse hazard rate $H\left(\cdot\right)$ as:

$$H\left(\overline{c}\right)\left(1-\eta\right)\left[\frac{u}{\sigma\alpha\left(1-\beta\right)} + \left(v_1 - v_2\right)\left(1-r^S\right)\right] = 1, \tag{4}$$

$$H\left(\overline{c}\right)\left(1-\eta\right)\left[\frac{v_1 - v_2}{m_2 - m_1}m_1 + \left(v_1 - v_2\right)\left(1-r^S\right)\right] = 1. \tag{5}$$

There is no equilibrium at $r^S = r^O = 1$ if the left hand side (LHS) of either (4) or (5) is then less than 1. So a sufficient condition for distortions is,

$$H\left(u + v_1\right)\left(1-\sigma\alpha\right)m_1 \min\left\{\frac{u}{\sigma\alpha\left(1-\beta\right)m_1}, \frac{v_1 - v_2}{m_2 - m_1}\right\} < 1, \tag{6}$$

Moreover, since the LHS of both (4) and (5) are decreasing in both r^O and r^S, this condition is also necessary, giving:

Proposition 3 *The search engine allocates traffic inefficiently, $r^S < 1$ and/or $r^O < 1$, if and only if condition (6) holds.*

The minimum expression in (6) is instructive: G has stronger incentives to distort organic search (reduce r^O below 1) than to distort sponsored search (reduce r^S below 1) if $\frac{u}{\sigma\alpha(1-\beta)m_1} < \frac{v_1-v_2}{m_2-m_1}$, and conversely if the inequality is reversed. These two terms represent the respective cost-benefit ratios from marginally distorting organic and sponsored searches from $r^S = r^O = 1$: distorting organic search reduces consumer surplus at the rate u (product trade values are fixed at v_1 since $r^S = 1$) while raising G's ad revenues at the rate $\sigma\alpha\left(1-\beta\right)m_1$, as search-based trades substitute for display-based trades; meanwhile, distorting instead sponsored search (reducing r^S) reduces consumer surplus at the rate $v_1 - v_2$ while raising the value of G's sponsored ads at the rate $m_2 - m_1$, both per product-searching consumer. To have r^S and r^O both interior in $(0,1)$ requires exact equality of their respective cost-benefit ratios, otherwise the first-order conditions cannot hold simultaneously. So generically, at most one will be interior and we neglect the measure zero range of parameters satisfying $\frac{u}{\sigma\alpha(1-\beta)m_1} = \frac{v_1-v_2}{m_2-m_1}$. By contrast, the solution for r^O or r^S is interior for a set of parameter values with a non-empty interior.[30] So we can conduct relevant comparative statics, but first we interpret the effects in terms of externalities.

Externalities on publishers. G's choices (r^S, r^O) generate both vertical and horizontal externalities on publishers, whose aggregate profits are $F\left(\overline{c}\right)\sigma\alpha e\left(r^O\right)m_1$: (i) G has a vertical externality

[30]The LHS of (4) and (5) are continuous functions of parameters and endogenous variables. For instance, if $\frac{v_1-v_2}{m_2-m_1} < \frac{u}{\sigma\alpha(1-\beta)m_1}$ and (5) holds at $r^S = r^O = 1$, a small parameter change that decreases the LHS of (5) induces r^S to fall strictly and no change in r^O.

because both r^S and r^O raise consumer participation, $F(\bar{c})$; (ii) G has a horizontal externality, because r^O raises the "effectiveness" of display advertising, $e\left(r^O\right)$. In other words, G's policy determines both the mass of consumers engaged in online search where they demand offline products, and how this demand translates into purchases, via either display or search advertising. In the limiting case of $\beta = 1$ with the baseline assumption of symmetric publishers, G cannot reduce the effectiveness of display advertising, and G only affects publishers vertically. But if $\beta < 1$, G also affects the degree of substitutability between display and search advertising. Concretely, G can accomplish business-stealing from publishers' competing display ad platforms by distorting organic traffic, because this reduces the effectiveness of publishers' websites as advertising outlets. The distortion hurts consumers as well as publishers, even if $r^S = 1$. So, in contrast to typical horizontal relations where competition is healthy, the horizontal externality results in lower consumer, and total, welfare.

Targeting technology improvements. We can characterize all possible effects of marginal changes in σ by considering in turn the cases where r^O and r^S are at a corner. First, suppose that r^S is at a 0 or 1 corner while $0 < r^O < 1$. Equation (4) reveals that an increase in σ affects the incentives to set r^O through four different channels. (i) if $r^S \neq 1$, $\bar{c}$ increases because display-mediated trades are always best-match while search-mediated trades are not; this relaxes the participation constraint, so the decreasing reverse hazard rate implies more distortion in r^O. (ii) η increases, which lowers the incentive to raise participation since a smaller fraction, $1-\eta$, conduct product searches; this lowers the opportunity cost of distorting organic search to raise $1-\eta$. Next, the bracketed expression on the LHS of (4), clearly falling with σ, is equal to $\frac{\partial \bar{c}}{\partial r^O} / \frac{\partial \eta}{\partial r^O}$, which reveals σ's last two effects. (iii) $\frac{\partial \eta}{\partial r^O}$ increases; organic distortion becomes a more effective tool for business-stealing. (iv) $\frac{\partial \bar{c}}{\partial r^O}$ increases; participation becomes more sensitive to r^O, which raises the cost of distorting organic search. Only this last effect (iv) has a positive sign and it is dominated by (iii). So, a higher value of σ induces the search engine to set a lower value of r^O. Second, suppose instead that r^O is at a 0 or 1 corner while $0 < r^S < 1$. Equation (5) reveals two influences of an increase in σ on r^S: (i) again, $\bar{c}$ increases, relaxing the participation constraint and implying more distortion in r^S. (ii) $\frac{\partial \bar{c}}{\partial r^S}$ falls, so participation becomes less sensitive to r^S.[31] Both effects induce the search engine to set a lower value of r^S.[32] Summarizing:

Proposition 4 *Targeting technology improvements induce the search engine to reduce reliability; increasing σ strictly reduces r^S and r^O whenever interior.*

So marginal changes in σ induce one of r^O and r^S to move in a common direction.

[31] The second term, $1-\eta$, can be expressed as $\frac{\partial \bar{c}}{\partial r^S} / (v_1 - v_2)$. Unlike in (ii) above, the fall in $1-\eta$ affects the costs and benefits of changing r^S in the same proportion.

[32] In all cases, σ has a positive direct effect on consumer participation, but this may be outweighed by the negative indirect effects of reduced r^O or r^S. σ's overall effect on η is similarly ambiguous.

The two instruments, r^O and r^S, are imperfect substitutes from G's perspective; G's profit function satisfies the standard definition of substitutability: $\frac{\partial^2 \Pi_G}{\partial r^S r^O} < 0$. These instruments interact through two channels. First, they both increase consumer participation. Second, a higher value of one reduces the sensitivity of consumer participation with respect to the other; this is because consumers are less harmed by sponsored search distortions when they more often find products via display ads and conversely, consumers are less harmed by organic distortions when sponsored searches lead to better product matches.

To conclude this section, we remark on particular parameter choices that specialize our model to two simpler models, each characterized by only one type of search distortion – organic and sponsored, respectively. For m_1 sufficiently close to m_2, the LHS of (5) is sufficiently large that G's optimal choice is always $r^S = 1$; obviously $r^S = 1$ also holds when m_1 exceeds m_2. That is, if the conflict of interest, as measured by $m_2 - m_1$, is sufficiently small or negative, G optimally sets $r^S = 1$ and only considers engaging in organic search bias, $r^O < 1$. Similarly, under the baseline assumption of publisher symmetry**,** as β approaches 1 the LHS of (4) approaches infinity and G's optimal choice is always $r^O = 1$. In other words, if the ability to disrupt display advertising, as measured by $1 - \beta$, is sufficiently small, G only considers product search bias, $r^S < 1$. These are special cases of our model so the results still apply, just simplified by fixing either r^S or r^O at unity.

4 The effect of integration with full monopolization

In this section, we examine the effects of integration in the baseline model, beginning with an idealized scenario that isolates the positive sides of integration. Here and in sections 5 and 6, we study integration into ad intermediation, deferring integration into publishing to section 7. As already noted, integration has both a vertical and a horizontal dimension. The vertical relation arises because the search engine contributes an input in the production of display-mediated trades by sending consumers to relevant content websites where publishers and ad intermediaries display ads. Integration can internalize this vertical externality and motivate greater search reliability as a way to attract more consumers. The integration also has horizontal features, since display advertising is an imperfect substitute of sponsored search advertising. By internalizing the horizontal externality, integration can remove the incentives to engage in the inefficient business-stealing described in the last section. As we show in later sections, our initial results are somewhat extreme, because there is no regulation to limit distortions nor to limit monopolization, and because publishers are perfectly symmetric and have no bargaining power when negotiating with G.

We consider a merger between G and one of the ad intermediaries and we refer to publishers that pay

G's ad intermediary to run their display advertising as "affiliates" (of G) and the rest as "non-affiliates" (of G). Given the absence of regulatory supervision, G can set different values of organic search reliability for affiliates and non-affiliates. We also need to explicitly describe the actions of intermediaries. So the baseline model changes in two ways. In the first stage, G now announces $\left(\mu, r^S, r_G^O, r_{NG}^O\right)$, where r_G^O and r_{NG}^O are the reliability of organic results to affiliated and non-affiliated websites, respectively. In between the first and second stages of the baseline model, each intermediary simultaneously announces its tariff T for publisher services, and then publishers respond simultaneously.

G's ad intermediary can handle an unlimited number of publishers, so it can capture the entire surplus from display advertising. In equilibrium, all other intermediaries offer $T = 0$, as was left implicit in the previous section, and G sets $r_{NG}^O = 0$ to extract all publisher rents by charging $T_G = \frac{F(\overline{c})}{N}\eta m_1$, which each publisher accepts. In this equilibrium, the average reliability of content search, r^O, satisfies $r^O = r_G^O$. Nothing changes beyond the transfer of rents to G. G's profits are now:[33]

$$\Pi^G = F\left(\overline{c}\right)\left[\left(1-\eta\right)M\left(r^S\right) + \eta m_1\right].$$

Relative to non-integration, G's per-consumer profits rise by the rents ηm_1 extracted from publishers. That is, G now gains from consumer participation through display as well as sponsored search auctions. So G has stronger incentives to attract consumer participation and this encourages increased reliability of both organic and sponsored search. In addition, since η increases with r^O, G can increase these new rents by raising the reliability of organic search, which makes display ads more effective. In consequence, r^S and r^O both increase weakly, as we now prove in detail. Mathematically, the first-order conditions for maximizing G's profits with respect to r^O and r^S, respectively, shift from conditions (4) and (5) to:

$$\frac{\partial \Pi^G}{\partial r^O} = f\left(\overline{c}\right)\frac{\partial \overline{c}}{\partial r^O}\left[\left(1-\eta\right)M\left(r^S\right) + \eta m_1\right] - F\left(\overline{c}\right)\frac{d\eta}{dr^O}\left[M\left(r^S\right) - m_1\right] = 0, \tag{7}$$

$$\frac{\partial \Pi^G}{\partial r^S} = f\left(\overline{c}\right)\frac{d\overline{c}}{dr^S}\left[\left(1-\eta\right)M\left(r^S\right) + \eta m_1\right] + F\left(\overline{c}\right)\left(1-\eta\right)\frac{dM\left(r^S\right)}{dr^S} = 0. \tag{8}$$

which can be rewritten as:

$$H\left(\overline{c}\right)\frac{\left(1-\eta\right)M\left(r^S\right) + \eta m_1}{M\left(r^S\right) - m_1}\left[\frac{u}{\sigma\alpha\left(1-\beta\right)} + \left(v_1 - v_2\right)\left(1 - r^S\right)\right] = 1, \tag{9}$$

$$H\left(\overline{c}\right)\left[\frac{v_1 - v_2}{m_2 - m_1}m_1 + \left(v_1 - v_2\right)\left(1 - r^S\right)\left(1-\eta\right)\right] = 1. \tag{10}$$

The LHS is higher for (9) than (10), for (9) than (4) and for (10) than (5), which indicate respectively, *(i)* $r^O \geq r^S$ with at most one being interior, *(ii)* given r^S, integration encourages more reliable organic

[33] Here, we assume G cannot influence the type or quantity of display advertising of publishers affiliating with G. See also sections 7 and 8.

search, and *(iii)* given r^O, integration encourages more reliable sponsored search. In fact, we can show that interaction between these two instruments does not change these insights. In particular, in the Appendix we prove:

Proposition 5 *Integration with full monopolization improves the reliability of the search engine and increases both consumer and total surplus; moving from no integration to full integration weakly raises r^S and r^O, and raises one or both strictly, unless initially at a corner solution (i.e., an extreme, 0 or 1, value for each of r^S and r^O).*

Under integration, G monopolizes the display advertising intermediation market and appropriates all publisher rents, which improves G's incentives to allocate traffic correctly, from the consumer and total surplus perspectives, for both types of queries. G internalizes both the vertical and horizontal externalities discussed above. The incentives to distort organic search are lower because the business-stealing effect disappears – G internalizes the horizontal externality. In addition, G takes into account the effect of higher consumer participation on publishers' rents – G internalizes the vertical externality from attracting consumers with higher reliability of both organic and sponsored search.[34]

This proposition ignores the distributional consequences of vertical integration. In particular, publishers get zero profits which would, in a model with costly content, affect the quality of online content, as we explain in section 9. Another potential drawback of integration lies in G's incentives to discriminate against publishers that do not deal with G's ad intermediaries or are otherwise less productive for G. This effect is absent in the above, extreme case of full monopolization where publishers are symmetric and all deal with G in equilibrium. In the next two sections, we illustrate how discrimination comes into play in a less extreme market structure.

5 Integration with partial monopolization

The assumptions in our baseline model, in particular constant returns to scale in the ad intermediation technology and the absence of regulation, result in full monopolization when G enters the ad intermediation market. That is, in equilibrium, all publishers affiliated with G's ad intermediary. So G's policy of discriminating against the publishers not dealing with G's ad intermediary did not translate into any discrimination among publishers *in equilibrium*. In a more realistic setting, publishers would be heterogeneous and G's integration with one ad intermediary would typically result in partial monopolization. So in this section, we start examining the consequences of partial monopolization of the

[34]The only remaining incentive to distort organic search derives from the fact that the integrated entity gets higher per consumer profits via search than via display ads – $M\left(r^S\right) \geq m_1$; cf., section 7. As with non-integration, improved targeting technologies increase remaining incentives to distort traffic.

display advertising market. Rather than model the exact obstacles that prevent full monopolization, we suppose that G's ad intermediary can handle the advertising business of at most a fraction γ of publishers.[35] We will show that integration can then lower welfare.

As in the previous section, we let G treat affiliated and non-affiliated websites differently, by setting r_G^O and r_{NG}^O, and again, G offers ad intermediation in exchange for a tariff, T_G, seeking to attract the maximal feasible fraction, γ, of publishers. We assume consumers do not know whether their favorite publisher will be affiliated with G when they decide on participation.

Clearly, G will find it optimal to send any diverted traffic to a publisher in the G network of affiliates.[36] So extending the effective visit notation, $e\left(r^O\right)$, from previous sections, by letting e_G and e_{NG} denote the average aggregate "effectiveness" of visits to publishers inside and outside the G network, respectively, we have,

$$e_G\left(r_G^O, r_{NG}^O\right) = r_G^O + \left(1 - r_G^O\right)\beta + \frac{1-\gamma}{\gamma}\left(1 - r_{NG}^O\right)\beta,$$

$$e_{NG}\left(r_{NG}^O\right) = r_{NG}^O,$$

assuming that indeed the fraction γ of publishers accept G's offer. The fraction of trades occurring on G's affiliated and non-affiliated publishers is then $\eta_G = \sigma\alpha e_G$ and $\eta_{NG} = \sigma\alpha e_{NG}$, respectively, so the overall fraction of trades occurring via display advertising, η, is given by:

$$\eta = \gamma\eta_G + (1-\gamma)\eta_{NG} = \sigma\alpha\left(\gamma e_G + (1-\gamma)\, e_{NG}\right),$$

while the average accuracy of organic search results, r^O, is:

$$r^O = \gamma r_G^O + (1-\gamma) r_{NG}^O.$$

Participation is still determined by (1) using these averages. Lemma 3 still holds, both for publishers affiliated with G and for the non-affiliated: type 1 merchants still bid a PPC of σm_1 in both types of publishers' auctions. It only remains to determine how G sets T_G in the continuation game: G's optimal choice is a tariff equal to a publisher's willingness to pay to be part of the G network. Expecting a fraction γ to affiliate with G, this value is:

$$T_G = \frac{F\left(\overline{c}\right)}{N}\left(\eta_G - \eta_{NG}\right) m_1.$$

[35] One possible interpretation of this exogenous constraint is that an increase in G's market share above γ might trigger an unwanted investigation by the regulatory agency.

[36] Distorting search for online goods away from the best-match publisher has the same effect on customer participation and merchants' willingness to pay for sponsored ads regardless of whether the *destination* publisher is affiliated with G; in equilibrium, all publishers have the same quantity and type of display ads. Thus, the destination of diverted traffic is irrelevant from the cost point of view, but not from the benefit point of view: G can charge its affiliated publishers a higher tariff if it sends them all the additional traffic diverted from non-affiliated sites.

G's expected profits from these affiliated publishers sum to $\gamma N T_G$ in total, giving overall expected profits:

$$\Pi^G = F(\overline{c}) \left[(1-\eta) M\left(r^S\right) + (\eta - \eta_{NG}) m_1 \right]$$

where we use the fact that $\eta_G - \eta_{NG} = \frac{\eta - \eta_{NG}}{\gamma}$. The effects of r_G^O and r^S on G's profits are analogous to those discussed in the case of full monopolization, equations (7) and (8).

In equilibrium, all publishers earn a rent equal to what they could get by refusing G's offer. So publishers jointly appropriate $\eta_{NG} m_1 = r_{NG}^O \alpha \sigma m_1$. Clearly, r_{NG}^O increases this publisher surplus. In addition, for any $r^S < 1$, r_{NG}^O reduces the total producer surplus, by increasing the share η of display-mediated trades, which contribute a lower margin, m_1, to producer surplus than the alternative of search-mediated trades with average margin, $M\left(r^S\right)$. Since G appropriates the producer minus the publisher surplus, G has an incentive to reduce r_{NG}^O for both these reasons. In the full monopolization case, G always minimizes r_{NG}^O at zero to extract all publisher rent. However, consumer participation is now increasing in r_{NG}^O, because organic distortions affecting non-affiliated publishers now affect a non-zero fraction, $1-\gamma$, of consumers in equilibrium. This moderates G's incentives to reduce r_{NG}^O and the optimal r_{NG}^O may be positive, but for exactly the same reason, any such distortions ($r_{NG}^O < 1$) now impose equilibrium inefficiency: they reduce consumer and social surplus. We now demonstrate how these harms can dominate the positive side of integration so that partial integration decreases overall social surplus, relative to non-integration.

As just explained, the main distinctive feature of partial integration is reflected in the first-order condition with respect to r_{NG}^O:[37]

$$\frac{\partial \Pi^G}{\partial r_{NG}^O} = f(\overline{c}) \frac{\partial \overline{c}}{\partial r_{NG}^O} \frac{\Pi^G}{F(\overline{c})} - F(\overline{c}) \frac{\partial \eta}{\partial r_{NG}^O} \left[M\left(r^S\right) - m_1 \right] - F(\overline{c}) \frac{\partial \eta_{NG}}{\partial r_{NG}^O} m_1. \tag{11}$$

To emphasize the role of the fraction of affiliated publishers, γ, we rewrite this as:

$$\frac{\partial \Pi^G}{\partial r_{NG}^O} \frac{1}{(1-\beta)\sigma\alpha} = (1-\gamma) \Psi\left(r_G^O, r_{NG}^O, r^S\right) - F(\overline{c}) \frac{m_1}{1-\beta}, \tag{12}$$

where we embed the first and second effects in $\Psi\left(r_G^O, r_{NG}^O, r^S\right)$ defined by:

$$\Psi\left(r_G^O, r_{NG}^O, r^S\right) = f(\overline{c}) \left(\frac{u}{(1-\beta)\sigma\alpha} + (v_1 - v_2)\left(1 - r^S\right) \right) \frac{\Pi^G}{F(\overline{c})} - F(\overline{c}) \left[M\left(r^S\right) - m_1 \right].$$

As γ increases towards 1, the first two effects, embedded in Ψ, converge to 0, but the third effect remains strictly negative in the limit. So for sufficiently large γ, G has incentives to distort traffic to non-affiliates; this distortion has little impact on consumer participation but a non-negligible impact on the rents that G can collect from its affiliated publishers.

[37] The three terms correspond to the three effects just described, now in inverse order: raising r_{NG}^O raises participation, lowers producer surplus and raises publisher surplus.

As with full monopolization in the previous section, partial integration leads G to internalize the horizontal and vertical externalities on the affiliated publishers. The vertical internalization is partial, because publishers retain some rents if $r^O_{NG} > 0$. The horizontal internalization is also partial, because it only applies to affiliated publishers. Nonetheless, these effects promote reliable search results for sponsored search and organic searches where an affiliate publisher is relevant; that is, r^O_G, r^S are higher. At the same time, partial integration exacerbates G's incentives to steal business from non-affiliated publishers, because distorting organic searches that should be directed to non-affiliated publishers not only substitutes search-mediated trades for display-mediated trades (as occurred with non-integration) but also transfers the remaining display-mediated trades from non-affiliated to affiliated publishers. This exacerbated business-stealing can make consumers worse off under partial than non-integration. For a simple illustration, suppose condition (6) is broken so that G sets $r^O = r^S = 1$ under non-integration. As just explained above, under partial integration with γ sufficiently close to 1, G sets $r^O_{NG} < 1$. Both consumer and total surplus are then lower than under non-integration. Summarizing:

Proposition 6 *There exists a region of parameter values for which vertical integration with partial monopolization reduces both consumer and total surplus.*

The illustration identifies the converse of (6) and γ close to 1 as sufficient conditions for integration to cause a welfare reduction. This might suggest relatively small welfare losses, given that only a fraction, $1 - \gamma$, of publishers face distortion. However, in the Appendix, we show that γ can be arbitrarily low and the consumer and total surplus losses occasioned by integration can be substantial.

6 Asymmetric publishers

In this section, we analyze how differences in publishers' effectiveness for display advertising affect search bias incentives and welfare outcomes. It is important to also allow for imperfections in the product search channel for offline trade. Otherwise, as shown above in the baseline model, publisher asymmetries are irrelevant to an integrated monopolist, because display advertising is redundant, exactly as in the first-best. As already noted, that redundancy was an artifact of the assumption of frictionless product search. Introducing product search imperfections gives the integrated monopolist a natural motive to distort organic search to raise the effectiveness of display advertising. While the baseline model only allowed for distortions that reduced display effectiveness (via $\beta < 1$), publisher heterogeneity in display effectiveness now makes display-enhancing distortions feasible too. We demonstrate that this new distortion incentive can make integrated monopoly worse for consumers and social welfare than non-integration.

We model heterogeneity in publisher ad effectiveness in the simplest possible way: a proportion ρ of

publishers are type H, characterized by a higher baseline ad effectiveness α_H than the rest, which are type L and have effectiveness $\alpha_L < \alpha_H$. We denote $\alpha = \rho\alpha_H + (1-\rho)\alpha_L$. We assume $\alpha_L < \beta\alpha_H$ to focus on the interesting, high asymmetry case where display advertising is always less effective on type L than type H publishers, even comparing best-match visitors on the type L publishers with worst-match visitors on type H publishers. To introduce imperfection in product search, we now assume that, while a proportion ϕ of participating consumers behave as in the baseline model, always able to search for products at no added cost after searching for content, the remaining proportion, $1-\phi$, can only conduct a content search. Publishers, intermediaries, merchants, and the search engine know this, but do not know which individual consumers have a viable option of product search. Consumers, by contrast, learn whether they can do a product search after choosing to participate and before responding to display ads during content search.[38] Notice that during their content search, a fraction $1-\phi$ of consumers are now willing to buy when faced with a display ad of the relevant type 2 product. But the other fraction ϕ would still wait to conduct a product search, so if ϕ is reasonably large, type 1 merchants still outbid type 2 merchants when competing for display ads. We assume $m_1 > (1-\phi)m_2$, to indeed ensure that display ads are all of type 1 products, as in the preceding analysis.[39]

G may gain by treating asymmetric publishers asymmetrically. So we distinguish the reliability of content search by whether the consumer is looking for a type H or type L publisher, denoting by r_H^O and r_L^O, respectively. Moreover, G can choose to divert customers to either of the two types of publishers, so we let $d_{a,b}$ represent the fraction of traffic diverted from type a publishers that is directed to type b publishers, for $a, b \in \{L, H\}$; of course, $d_{aH} + d_{aL} = 1$ for each $a \in \{L, H\}$. Now η becomes,

$$\begin{aligned}\eta = \sigma\rho\left[r_H^O\alpha_H + (1-r_H^O)\beta\left(d_{HL}\alpha_L + (1-d_{HL})\alpha_H\right)\right] \\ + \sigma(1-\rho)\left[r_L^O\alpha_L + (1-r_L^O)\beta\left(d_{LH}\alpha_H + (1-d_{LH})\alpha_L\right)\right].\end{aligned} \tag{13}$$

The average reliability of content search is $r^O = \rho r_H^O + (1-\rho)r_L^O$, and participation is now given by,

$$\overline{c} = r^O u + \eta v_1 + \phi\left(1-\eta\right)\left(v_1 r^S + v_2\left(1-r^S\right)\right).$$

We begin with the case of non-integration. G's profits are,

$$\Pi^G = \phi F\left(\overline{c}\right)\left(1-\eta\right)M\left(r^S\right).$$

For a given level of sponsored search reliability r^S, if one can fix the fraction of display-mediated trades, η, G's profits are increasing in r^O since this raises participation $\overline{c}$ without changing G's profit per

[38] One interpretation is that, after deciding to participate, a fraction ϕ of consumers will discover a product need and form a plan to search for and buy it, but the other fraction, $1-\phi$, are unaware of their latent product demand and only consider buying offline products if exposed to an ad.

[39] This inequality is unnecessary if consumers always discover how to conduct a product search upon seeing a display ad of the relevant product, of either type.

consumer. Since η is increasing in r_H^O and decreasing in d_{HL}, if these two values are interior, one can raise d_{HL} and r_H^O such that η remains unchanged but r^O indeed increases. Thus:

Lemma 4 *Under non-integration, if $r_H^O < 1$, then $d_{HL} = 1$.*

The intuition is simple. Under separation, G benefits from reducing offline trade mediated through publisher display ads, but wants to let consumers consume their preferred content as far as possible. Distorting traffic from type H to type L publishers maximizes the reduction in display-mediated trade for a given reduction in the reliability of content search.

To evaluate the possibility of an undistorted equilibrium, we suppose $r_L^O = r^S = 1$ and focus on the incentives over r_H^O. In this case, $r^O = \rho r_H^O + (1-\rho)$. Considering the first-order conditions for G's profit maximization evaluated at $r^S = r_H^O = 1$ and $\eta = \sigma\alpha$, we can write the conditions for an equilibrium with no distortions as:

$$H\left(\overline{c}\right)\left(1-\sigma\alpha\right)\left(\frac{u}{\sigma\left[\alpha_L-\beta\left(d_{LH}\alpha_H+(1-d_{LH})\alpha_L\right)\right]}+v_1(1-\phi)\right)\geq 1,$$

$$H\left(\overline{c}\right)\left(1-\sigma\alpha\right)\left(\frac{u}{\sigma\left(\alpha_H-\beta\alpha_L\right)}+v_1(1-\phi)\right)\geq 1, \tag{14}$$

$$\phi H\left(\overline{c}\right)\left(1-\sigma\alpha\right)\frac{v_1-v_2}{m_2-m_1}m_1\geq 1, \tag{15}$$

where $\overline{c} = u + \left(\sigma\alpha(1-\phi)+\phi\right)v_1$. The first inequality is implied by the second. Note that for $\phi = 1$ and $\alpha_H = \alpha_L$, this replicates the conditions in section 3. As in that section, these conditions are also sufficient; the LHS of both (14) and (15) are decreasing both in r^S and in r_H^O.

We now turn to the case of integration where G monopolizes the ad intermediation market. As in section 5, we seek to demonstrate parameter values for which G will distort under integration with full monopolization, but not under separation. So suppose that G sets $r^S = 1 = r_H^O$. In this case, $r^O = \rho + (1-\rho)r_L^O$, $\overline{c} = r^O u + \phi v_1 + \eta(1-\phi)v_1$ and

$$\Pi^G = m_1 F\left(\overline{c}\right)\left[\phi+\eta(1-\phi)\right].$$

Clearly, Π^G is now increasing in η as well as $\overline{c}$, which is itself increasing in η, for any given value of r^O. (G has no cost from increasing η when integrated, given G's search ads are selling type 1 products just like the display ads.) It follows immediately that $d_{LH} = 1$ is optimal since d_{LH} increases η. Since $\frac{\partial\eta}{\partial r_L^O} = \sigma(1-\rho)\left(\alpha_L-\beta\alpha_H\right)$ is negative by our assumption of substantial publisher heterogeneity, G now has a motive to decrease r_L^O, diverting search to more ad-effective publishers. Of course, there is a tradeoff, because search accuracy raises participation. These effects are captured in the first-order derivative which we evaluate at $r_L^O = 1$,

$$\frac{\partial\Pi^G}{\partial r_L^O}\frac{1}{m_1} = f\left(\overline{c}\right)\frac{\partial\overline{c}}{\partial r_L^O}\left[\phi+\eta(1-\phi)\right]+F\left(\overline{c}\right)\frac{\partial\eta}{\partial r_L^O}(1-\phi), \tag{16}$$

Notice that if $\phi = 1$, the second term representing the increase in display effectiveness is nullified since display is redundant, as explained earlier.[40] Now a sufficient condition for the existence of distortions under vertical integration is that,

$$H(\overline{c})\frac{\sigma\alpha+\phi(1-\sigma\alpha)}{1-\phi}\left(\frac{u}{\sigma(\beta\alpha_H-\alpha_L)}-v_1(1-\phi)\right)<1, \tag{17}$$

evaluated at $\overline{c} = u+(\sigma\alpha+\phi(1-\sigma\alpha))v_1$ as in (14) and (15). It is a simple exercise to see that condition (17) is compatible with (14) and (15), and so integration may lead to an increase in distortions in organic search. This increased distortion necessarily reduces both consumer and total surplus when $u > (m_1+v_1)(1-\phi)\sigma(\beta\alpha_H-\alpha_L)$. A consumer interested in an L type publisher may be attracted by a display ad with higher probability if sent to an H type publisher. The probability increase is $\sigma(\beta\alpha_H-\alpha_L)$, and with probability $1-\phi$, this is the consumer's only chance for consuming an offline product. However, given the inequality, this potential gain in surplus does not compensate the consumer's direct utility loss from not consuming her preferred online content. This sufficient condition for distortions to result in consumer and total surplus losses is also compatible with (17), (14), and (15).[41] Thus, we conclude,

Proposition 7 *Integration with full monopolization can reduce consumer and total surplus.*

As in the case of symmetric publishers, integration with full monopolization induces G to internalize the vertical and horizontal externalities imposed on publishers. The difference is that G's internalization of the vertical externality may no longer be in the interest of consumers. Indeed, under the assumptions of this section, aggregate publisher revenue is larger if traffic is distorted from publishers with low to high ad effectiveness. Moreover, G cares about this display ad revenue, which is not redundant given the friction, $\phi < 1$, in product search and search-mediated advertising. So G has a new incentive to distort traffic, this time from less to more ad-effective publishers.

We could have also considered the case where G only deals with type H publishers, perhaps because it cannot price-discriminate among publishers and prefers to set a tariff too high for type L publishers to accept. In this case, both the incentives to distort traffic away from non-affiliates and away from

[40] As is already implicit, publisher asymmetry is also crucial: the key factor in the display ad effectiveness derivative, $\alpha_L - \beta\alpha_H$, is obviously negative if $\alpha_L = \alpha_H$. The distortion arises because with $\phi < 1$, display is not redundant and publisher asymmetry gives G an instrument that increases η: G can direct consumers whose preferred publisher is type L to a type H publisher. With symmetry, the effect of r^O on Π^G, for $\phi = 1$ represented by (7), is always positive if $r^S = 1$. Introducing $\phi < 1$ to the symmetric case, $\frac{\partial\Pi^G}{\partial r^O}$ is given by (16) with r^O substituting for r_L^O. Since participation and profits per consumer would then both increase with η and participation and η would still increase in r^O, $\frac{\partial\Pi^G}{\partial r^O}$ would still be positive.

[41] For $u = (m_1+v_1)(1-\phi)\sigma(\beta\alpha_H-\alpha_L)$ and $\beta = 1$, the parenthesis in (17) takes the value $m_1(1-\phi)$, and the parenthesis in (14) takes the value $(2v_1+m_1)(1-\phi)$. So for v_1 sufficiently large, the left hand side of (17) is smaller than the left hand side of (14), and the value of $H(\overline{c})$ can be adjusted so that the two expressions lie on the corresponding sides of 1. Finally, a small enough value of $m_2 - m_1$ guarantees that (15) holds.

low ad-effective publishers would coincide. The partial internalization of both vertical and horizontal externalities would then combine against the interest of consumers, reinforcing the possibility that integration lowers consumer and total surplus.

7 Integration with publishers

So far, we have considered integration of the search engine with an intermediary from the display advertising market. Integration with publishers results in similar incentives for G. Indeed, if the integrated entity did not modify how it handled display advertising, sections 4 and 5 would continue to describe the effects of integration on G's policies. However, integration with publishers may have slightly different consequences, since this integration is likely to facilitate manipulation of the supply of advertising and the coordination of pricing strategies. In fact, there is no role for price coordination in our stylized model, since prices already extracted all merchant rents under non-integration. But by manipulating the supply of different types of advertising, an integrated entity might raise its profits beyond our section 4 and 5 predictions. To show this, we first consider the simplest and extreme case, where the search engine owns all publishers, but has them set display advertising as in section 4. For the parameter values that gave $r^S = 0$ and $r^O < 1$ in section 4, this would imply the profits derived there as,

$$\Pi^G = H\left(\bar{c}\right)\left[\left(1-\eta\right)m_2 + \eta m_1\right],$$

with $\bar{c} = r^O u + \eta v_1 + (1-\eta)\, v_2$. The reason for distorting r^O was that display ads, being of type 1, restricted G's ability to maximize type 2 offline trades; the combined entity distorted organic results to transfer advertiser attention to its search platform where it could better exploit that attention. However, now the integrated entity could simply choose to eliminate display advertising from publishers' websites, inducing all consumers to conduct product searches and removing the motive for distorting organic search. G would then set $r^O = 1$ and could replicate its previous level of per-consumer profits from display and search ads by setting r^S equal to the prior level of η, which also replicates consumers' offline trade distribution. Since the increase in r^O raises consumer participation, the integrated entity gains strictly by removing display advertising. Alternatively, it could restrict display advertising to high margin, that is type 2, merchants and again adjust its results strategies as just described. In the baseline setting, this is equivalent, since no one buys via type 2 display ads given any $r^S > 0$. But in the model with $\phi < 1$, the alternative of restricting a fraction of display advertising to type 2 merchants is strictly preferable to shutting down display ads, because the display ad channel is not redundant for the fraction $1-\phi$ of participating consumers who cannot be reached by search advertising.[42]

[42] In either case, with *full* integration into publishing, G would have no incentive to distort organic search.

Similarly, when the search engine owns a fraction γ of publishers and G's publishers maintain their display ads, then borrowing from section 5, there is a region of parameter values such that $r_G^O < 1$ and $r^S = 0$. A more profitable policy includes blocking type 1 display advertising in their own publishers, setting $r_G^O = 1$, leaving r_{NG}^O fixed, and setting $r^S = \frac{\gamma\eta_G}{1-(1-\gamma)\eta_{NG}}$. With this policy, the integrated entity makes the same profits per consumer and induces higher consumer participation. Again, simply shutting down display ads is an optimal strategy if $\phi = 1$, but more generally, G would only want to restrict against low margin display ads. Summarizing,

Proposition 8 *Integration with publishers has essentially the same effects as integration into ad intermediation but may differ by reducing the supply of display advertising or restricting display to high margin products.*

This minor difference would disappear if a G owned ad intermediary were also able to exert influence over display advertising.[43] For instance, G might exert influence by committing to divert traffic away from those publishers that refuse to adjust the quantity and content of their display ads to G's request; see also section 8 on publisher discrimination. Finally, notice that an analogy of the endogenous affiliation process described for a G ad intermediary in section 4 could, in principle, lead G to monopolize the market for publishing, diverting traffic away from publishers that reject a G buyout. However, this seems less plausible than the, already extreme, case of a monopolizing intermediary, for two reasons. First, ad contracts are regularly renewed, facilitating repeated game effects compared to ownership transfers. Second, publishers are more visible to consumers and plurality of ownership is recognized by regulators as a fundamental value in the media context.

8 Discussion

The model presented in this paper is stylized and parsimonious. In Section 2, we devoted very little space to explain and motivate our modeling choices, so we now go back to the primitives of the model to justify some of the assumptions and discuss the robustness of the main results. We show how several features that might appear simplistic or far-fetched are actually equivalent to richer and more realistic representations which readily generalize to fit a broader range of possible outcomes.

The timing of search. In the baseline model, people always enter their queries for online goods before those for offline goods. Our results on organic search bias only require that content searches

[43]The net consequence of this difference for consumers is ambiguous. With monopolizing publisher integration and publisher symmetry, organic search is always undistorted, but offline trades may be more distorted owing to the loss of type 1 display ads. With partial monopolization, publisher integration again may distort offline trades more, but distorts organic search less, than for comparable ad intermediary integration where business-stealing from non-affiliate publishers facilitates rent extraction from affiliated publishers.

sometimes result in display-mediated product trades that reduce people's need to product search. But the exact timing affects the intensity of the incentives for organic search distortions.[44] Typically, consumers will conduct their searches, for content or products, as needs come to mind. To capture this dynamic process in a simple static model, we could suppose each consumer conducts her searches in each of the two possible orders with positive probability, or even that some consumers only engage in one type of search as in section 6. Extending the model in this way scales down the interaction between the markets for display and search advertising, but the qualitative conclusions remain unchanged.

In the baseline model, consumers only conducted one content search and up to one product search. This is optimal given the assumptions of unit demands and G's ability to determine exactly what consumers seek, which we discuss below.[45] As noted in footnote 5, people spend significantly more time searching for content than products. The effect of seeking many units of content can be captured by an increase in u and α. This scales the display and search interaction back upwards.

Types of search result. An important simplification in the model is our sharp distinction between content searches, generating only organic results, and product searches generating only sponsored results. The relative intensity of sponsored relative to organic links in the list of search results is higher when the search objective is to find a product to buy – that is, a product or transactional search rather than a content search. Indeed, while product queries generate many sponsored results, content queries often generate none at all. But our depiction is clearly extreme. So we now explain why such a relationship should exist, as well as why it is less extreme in reality than in our model.

Product searches and sponsored results. For product searches, it is intuitive that profit-maximization leads G to withhold useful organic links: G cannot expect merchants to pay much for sponsored links if G already gives them useful organic links for free. In our model, G can simultaneously provide consumers any given quality of product matches *and* extract the full rent from the corresponding merchants by providing only sponsored links. If instead G provided organic links to relevant merchants independent of their bids in sponsored search auctions, those merchants would get a rent. Clearly, that would leave G worse off.

Of course, G suffers no loss in providing organic results that consumers know to be dominated by the sponsored results, because consumers would rationally never click on such organic links. We can use this fact to begin to understand why product searches generally provide organic, as well as sponsored, results in the more realistic environment where consumers only observe search reliability imperfectly. As discussed further below, G may then use explicit promises and avoid blatant bias, try to convince

[44] The incentives to bias sponsored search are not directly affected.

[45] For any positive r^S or r^O, it might seem that consumers would already wish to repeat search. In fact, there is no such incentive since G could provide the same results on each repeat search.

some consumers that organic results are never distorted. If G provided no organic results to certain searches, this would make it blatantly clear that G distorts its organic results so consumer participation would fall as trust in the promise becomes untenable. Similarly, regulators facing a high burden of proof may be able to punish when bias is so blatant.[46] So G would always provide some organic results and avoid other blatant distortions.

In fact, in the baseline model, G suffers no loss in providing a list of equally useful top organic and sponsored results, both of which would feature only winners of the sponsored auction.[47] Consumers would then be indifferent between clicking on the sponsored or the top organic results. Merchants would not pay directly when their organic results are clicked, but the equilibrium sponsored click price would rise.[48] Notice that merchants are then paying indirectly for high organic rankings. This would contravene standard claims that organic results are entirely unaffected by payments – see footnote 20 of Yang and Ghose (2010) for an official Google declaration. But the result is consistent with merchant complaints of suddenly losing their high organic rankings after they stopped sponsoring search.

Blake, Nosko and Tadelis (2013) present evidence from a field experiment in which eBay suspended its sponsored advertising for all queries that included its brandname among the keywords.[49] They found minimal reductions in traffic to eBay. In their data, searchers typically switched to clicking on eBay links among the organic results instead. This goes against the idea that G may remove organic links to merchants that stop bidding on sponsored links, but this may be specific to the fame of eBay. Indeed, in a related field experiment by a large national retailer which suspended search ads also for non-brandname keywords, Yang and Ghose (2010) found notably larger traffic reductions than in any of Blake et al.'s (2013) field experiments.[50] As in the above discussion of blatant distortions, there are strong reasons to expect greater reliability for navigational searches and searches where famous companies are clearly relevant. A search with a specific company, such as eBay, among the keywords is usually a navigational search; the consumer with this specific merchant in mind will see the failure to include that merchant high among the results as a blatant distortion. In addition, the relevant merchants might successfully

[46] Recent FTC and EU investigations into Google's search results centered on bias towards Google-owned publishers where evidence can be sharpest; the FTC found the evidence inconclusive; the EU is negotiating remedies.

[47] The organic list could be completed with a randomly ordered set of other merchants and internet sites, including enough irrelevant ones to put searchers off trying these links.

[48] For instance, if there is one sponsored and one top organic result and a fraction λ of consumers click on the sponsored result, merchants' willingness to pay in the sponsored auction would be $\frac{m_k}{\lambda}$, since the auction winner only pays per click on the sponsored link but now also receives traffic proportional to $1-\lambda$ from being the top organic result. The rest of the analysis would be entirely unaffected.

[49] Blake et al. (2013) also report on a field experiment where eBay suspended its sponsored advertising altogether for randomized geographic localities. Again, they found that eBay lost little traffic.

[50] Yang and Ghose (2010) explain their data by hypothesizing that sponsored and organic links are complementary in the generation of clicks through to a given merchant. This could arise if consumers believe a merchant link is more likely useful when the merchant appears high among *both* sponsored and organic results. In our framework with perfect commitment by G, there is no room for such an effect, but we generalize below.

denounce such distortions to regulators.

More generally, participation is more sensitive to salient search distortions than to less observable distortions. We now briefly discuss reputation as a precursor to better understanding the role of organic results.

Reputation and commitment. In the model, G sets μ, r^S and r^O and then consumers observe this before deciding whether or not to participate. Implicitly, this assumes G can commit perfectly to any distortion strategy, or reliabilities r^O and r^S, that it wishes to adopt. The commitment assumption captures in a static model the idea that, over time, G can build a reputation for reliable search results. A key constraint on reputation building is the difficulty consumers have in observing reliability. Perfect observation would require consumers to know the quality of the results that G *could* have provided, as well as seeing those it actually does provide. This extreme case is implausible, but consumers can certainly evaluate their own experiences and they may learn from each others' experiences. So some commitment is feasible. At the same time, the reality of imperfect commitment is useful for clarifying the joint participation constraint and the role of organic results.

Content searches. Publishers should, in principle, be willing to pay for traffic since visits (that lead to consumption of their content) raise their display ad revenues. So it is not obvious why publishers rarely sponsor links on content search queries. One reason why sponsored results might fail to dominate is that publishers may have negligible effective willingness to pay on each individual query, owing to transaction costs. Having to constantly adapt bidding strategies, over a myriad of potentially relevant keywords, to changes in content and query patterns can be very costly for publishers with highly dynamic content, such as news websites. This dissuades publishers from bidding.

Sponsored results are obviously influenced by merchants' willingness to pay. This is explicit in their very name.[51] By contrast, search engines proclaim that their organic results cannot be influenced by payments from anyone, calling these results variously, "organic, natural, unsponsored, unbiased" to emphasize this commitment.[52] So, on the face of it, these results should be the most reliable. To question the organic promise, a consumer must not only be skeptical over the explicit promises to not distort, but also sophisticated enough to notice and take account of the indirect strategic motives for distorting organic results, which can be quite subtle. In brief, G's possible interest in distorting sponsored results is explicit and salient, while the motives for distorting organic results are less salient and, for the most

[51] Indeed, Greenstein (2013) relates how Google began clearly identifying its sponsored ads as such (with a yellow background up top and to the right), separated from the "untainted" organic results (in white) precisely because of the founders' insight that advertising introduces a clear source of bias.

[52] As an exception to the rule, Marissa Mayer, while a Vice President at Google, recognized a bias towards own content sites like Google Finance: "*To the degree that we host content, we ultimately have a monetary incentive to drive people to those pages if those pages have ads on [them].*"

part, hotly denied and hard to prove.

This motivates the possibility, discussed above, that some consumers may trust in G's word – the promise of unbiased organic results – unless presented with strong evidence to the contrary. So long as G avoids biases that clearly belie the commitment, organic results then serve as an indirect commitment to relatively reliable results. More generally, consumers may just believe that the incentives to bias sponsored results are higher than for organic results for some class of searches, implying greater confidence in organic results. Empirical evidence suggests this is so for most content searches, though not necessarily for product searches (just as in our model), but why should organic results be particularly important for content searches?

This argument that organic results serve as an indirect commitment against search bias is relevant for queries where G has serious difficulties in committing directly to its desired reliability. Most online content sites offer informational goods which are relatively difficult to inspect and evaluate. So commitment is likely to be a greater problem there.[53] Providing organic results for these content searchers is a natural way to attract participants from whom G makes money via sponsorship of product searches.

Finally, we briefly explain why it is rare to see sponsored search advertising by merchants on content queries instead of organic results. First, such ads are likely to be ineffective, since it is more difficult to distract consumers into product consumption, while they are actively searching for content; in our model, ads for products only distract consumers once they have found satisfactory content and begun to consume it. Second, even if consumers seeking content *did* occasionally get distracted into clicking on sponsored links to merchants and *did* buy the product despite poor targeting, the negative participation effects of their frustration at having to keep on searching or abandon their content search is likely to outweigh any small advertising gains.

Joint participation constraint. We modeled search participation as a single decision, for both product and content searches, by each consumer. This simplification captures the fact that high quality results in either type of search tend to spill over into improving G's *overall* popularity or reputation as a reliable search engine. This fits the history of most general search engines: they began building a reputation for reliability, usually with just organic results, and started to make money from these participants by introducing sponsored results. One explanation for this spill-over is that consumers tend to develop a habit of using a fixed engine, rather than adapting each search to the specific search need of the moment. An independent but complementary explanation builds on the observability problems discussed just above. Given the many difficulties mentioned above for evaluating search

[53] Intuitively, reputation and credibility are key concerns in content search. Consumers are more willing to trust apparently independent recommendations, such as organic results, and be suspicious of publishers that need and wish to pay to reach them.

reliability, consumers tend to learn about search from friends, from media reports and from their own experience. Such learning tends to be coarse, rather than fully contingent on each type of search, because communication is limited and memory and aggregative skills may be limited too. A thorough analysis is beyond the scope of this paper, but notice that nothing in our analysis required consumers to observe more than the average gross benefit from searching on G – that is, the weighted average of r^O and r^S given by (1). Moreover, in the scenario where consumers indeed only observe this average benefit and are sufficiently averse to uncertainty over how this average breaks down into r^O and r^S, introducing separate costs of participating in content and product searches would have no impact on our analysis.

Singleton results list and ambiguous queries. We have assumed that G offers only one result, both for content and product searches; that is, the list of results, organic or sponsored, contains only one relevant item. This is in fact optimal since G knows all the relevant information about consumers' needs, and consumers have a unit demand for both online and offline goods.

In reality, most searches generate multiple results. This makes sense for consumers seeking multiple goods, but only if some of the goods demanded are sufficiently similar for a single query to identify them together. A more important reason for multiple results is that queries are often ambiguous, with consumers retaining private information or ability to evaluate results.[54] In the model, G was able to perfectly interpret all queries, in that G could identify the best-match content or products for any given query. With imperfect interpretation, G typically gains by offering multiple results (longer lists), because consumers can then inspect these results until they find a good that satisfies their need sufficiently well. This can improve matching because consumers tend to have private information or inspection ability. Consumers retain private information whenever they type in ambiguous queries to save on the costs of forethought and typing involved in precisely describing exact search preferences. These cost savings usually outweigh the costs of having to skim over multiple snippets or search links.[55] This plausible extension can readily accomodate long results lists. All our insights would still apply.

Conflicting interests betweeen consumers and merchants – more general consumer preferences. The driving force behind the distortion in sponsored results in our model is the potential conflict of interest between G, which can cash on merchants' margins, and consumers, whose net utility from purchases may not be aligned with these margins. For expositional clarity, we adopted an extreme misalignment, but the model readily accomodates more general specifications. For instance, instead of assuming that the consumer's best-match and second-best products always have low and high margins

[54]While G needs information on merchants' types to set $r^S > 0$, G does not actually need to be able to identify merchant categories and their relevance to different queries, because relevant merchants self-select in the keyword auctions. Nonetheless, ambiguous queries may make it hard for merchants to evaluate their consumer relevance and this has the same consequences as those we describe here.

[55]G may additionally benefit by learning to refine its interpretration skills over time.

(that is, are type 1 and type 2), respectively, we could allow any distribution over product type pairs. In this more general formulation, the sponsored auction will often present no conflict of interest: for queries where the consumer's best and second-best products are both of type 1 or both of type 2, the two merchant types would bid the same and G would always pick a best-match merchant as auction winner; similarly, when the best-match product is type 2, G would always let a best-match merchant win. These possibilities dilute G's incentives to distort sponsored search, but introduce no new qualitative effects. For example, in the special case where the pair of products are never misaligned, there is no conflict of interest in sponsored search and the analysis is equivalent to the case with $m_1 = m_2$ which we treated in section 3 as a special case of the general model; there, G's incentives for traffic distortions are limited to bias in organic results.[56]

In particular, while in the baseline model, the type 1 and type 2 merchants are vertically differentiated with the former preferred by all consumers, we can readily treat pure horizontal differentiation of merchants. Suppose that each consumer's preferred product is equally likely to be of either type. Here neither type has higher quality – it is simply that one or other type happens to be more relevant to that consumer's product search. In this case, the intensity of conflict of interest in sponsored search is simply half that in the baseline case, which was $\mathbb{E}\left((m_1 - m_2)(v_1 - v_2)\right)$.

Conflicting interests – endogenous merchant margins. A simple example with endogenous product pricing demonstrates one reason to expect a conflict of interest between consumers and merchants. Suppose consumers gain the same gross (of price) utility from each relevant merchant's product, but each merchant also sells to a local market of offline consumers that is more elastic in the case of type 1 than type 2 merchants. Then, if regulatory or reputational motives prevent merchants from price discriminating between their offline and online consumers, the type 1 merchants will set lower prices and have lower margins. This immediately implies a higher net value for consumers from the type 1 merchant, since quality is fixed and price is lower. As always, consumers and merchants have conflicting interest over price. In fact, $v_1 - v_2 = m_2 - m_1$ in this case.

Compulsive consumers. In the baseline model, the role of advertising is purely informative: consumers are fully rational and advertising merely enables consumers to locate merchants – advertising neither persuades nor tempts people to consume, nor does it complement consumption. This is the most common view of advertising in economic models and Blake et al. (2013) provide supportive evidence for this view in the case of search advertising. But we now consider the possibility that display advertising is persuasive as well as informative. This opens the door to more negative views of the welfare implications

[56] In mathematical terms, sponsored search is only distorted if sometimes $(m_1 - m_2)(v_1 - v_2) < 0$. If, in addition, G cannot observe consumer preferences perfectly (see above on ambiguous queries and interpretation problems), then $v_1 - v_2$ should be replaced by G's best estimate of it.

of advertising. The analysis readily extends, with few changes.[57]

Concretely, we now consider compulsive consumers who always consume either type of relevant product when tempted by a display ad, though rational consumers would instead wait to conduct a product search if the displayed ad is type 2. In this scenario, type 2 instead of type 1 merchants buy all the display ad slots, but the main tradeoffs for G are very similar. The main novelty is that compulsive consumers become tempted by type 2 display ads against their better interest (given $\phi = 1$), so that $\bar{c}$ is now decreasing in $\alpha\sigma$. Also compulsives might actually prefer organic search to be distorted. Obviously G would then distort organic search fully since there is no participation-profit tradeoff, so we consider the opposite case; a sufficient condition for consumers to prefer undistorted search is that $u > (v_1 - v_2)\,\sigma\alpha\,(1-\beta)$. In this case, the qualitative tradeoffs for search distortions are essentially the same as with rational, self-controlled consumers. Inducing participation is more difficult, since consumers anticipate suffering from the temptation of type 2 display ads, but for the same reason, distorting organic search has a lower participation cost.

Auctions and alternatives. Throughout the paper, we modeled specific scoring auctions determining prices in the form of PPC, price per click, both for display and search advertising. We now discuss alternatives and generalizations.

The scoring auction. The weighted position auction for determining PPCs seeks to capture, in a simplified framework, the mechanism that Google claims to use in reality. The outcome of the auction depends not only on merchants' bids but also on their quality scores, which capture the relevance of merchant products to consumer demands, as inferred from queries, as well as factors such as website quality. In so doing, Google recognizes the possibility of conflicting interests between consumers and merchants that we have described. Consumers would like G to position the producers of the type 1 good as the top sponsored result, but the producers of the type 2 good have a higher willingness to pay for this slot. The choice of μ and r^S reflects G's compromise between these two objectives.[58] Google's claims would suggest that $r^S = 1$, with μ purely serving to prevent type 2 merchants from winning, but the theory and the insight of Brin and Page (1998) cited in footnote 7 suggest that this may not be the case.[59]

Alternatives to auctions and non-necessity of the perfect competition assumption. In

[57] Advertising would have more negative welfare implications than we entertain here if advertising can even persuade consumers to consume net negative value goods with $m + v < 0$.

[58] In a previous version of this paper, μ also served to increase competition between merchants of different products, but that role disappears here with the two merchants per product competing for ad slots; that role for μ also disappears with symmetric information and posted offers as explained below.

[59] In the baseline model, consumers never do more than one product search, nor visit more than one publisher. But in general, G and publishers also provide the identity of each consumer visitor to merchants so that merchants know they are not paying multiple times for the same person. This is also necessary if merchants pay per impression (PPI) and are therefore unable to verify the number of impressions to unique visitors without some form of consumer identification.

our setup, setting a stochastic μ such that $\mu > \frac{m_1}{m_2}$ with probability r^S and $\mu < \frac{m_1}{m_2}$ with the converse probability generates an outcome equivalent to the auction design. Indeed, the second-price auction is equivalent to a mechanism where G simply sets take-it-or-leave-it offers to type 1 and type 2 merchants with probabilities r^S and $1 - r^S$, respectively. Notice that this posted offers alternative works equally well in the absence of competition between merchants of each product.[60] So this competition is not needed for our results.[61] Equally, the display auctions could be replaced by posted offers. However, in environments with publishers and G uncertain about merchant margins or G uncertain over consumer values, auctions can yield higher profits. Moreover, a deterministic value of μ generically determines merchants' success probabilities without need for tie-breaking.

Publishers optimally forego the possibility of using a weighted display ad auction, like μ, because merchants and publishers have a common interest given that only type 1 products ever sell. If consumers were compulsive, publishers would again share merchant preferences over display ad allocation, assuming sufficiently many publishers that each one essentially neglects the impact of display ads on consumer participation.

Display advertising slots and their effectiveness. Our derivation above for singleton sponsored results does not apply immediately to display advertising, because the probability of successfully distracting a consumer might vary with the mix of display ads, but notice that nothing would change if each publisher n were now able to subdivide its display ad space and sell the smaller slots to merchants, with each fraction of space providing that fraction of the distraction probability α. The references in footnote 9, especially Wilbur (2008), point to factors that may generate asymmetry in α across content websites.[62] To explain biased organic search in the case of fully symmetric content sites, if there is either no integration at all or fully monopolizing integration, the $\beta < 1$ assumption is a necessary condition.

Merchant heterogeneity in CTR or CR. The parameter α might also vary by merchant. We abstracted from such asymmetries, but they are analogous to Athey and Ellison (2011) and Chen and He's (2011) analyses of sponsored search when merchants vary in their click-through rates (CTR) and conversion rates (CR). These analyses are essentially compatible with ours. In our model of product search, both CTR and CR equal unity for any relevant product. Search engines claim to take account of these factors in weighting sponsorship bids. In particular, Google was the first to introduce click-weighted auctions in 2003, successfully reducing the prevalence of ads with low CTR, such as mobile

[60] Also all these designs for selling advertising are optimal in our setting as they extract full merchant rent.

[61] Of course, sponsored search bias does require two types of merchant to be interested in a common search query.

[62] Notice that our result of content search distortion in the case of monopolizing integration in this asymmetric setting is analogous to the monopoly case of an advertiser-funded media outlet in Ellman and Germano (2009): the monopoly outlet chooses news that makes advertising more effective and is akin to G sending consumers to a website with this kind of news as its content.

phone ringtone and porn-site ads. We could extend our model along the lines of these prior papers. They demonstrated positive self-selection so there would be no added conflict of interest, at least for the case with a single sponsored link.[63]

PPC versus price per impression, PPI. Throughout the paper, we assumed sale of ad space based on price per click PPC. The CTR for sponsored links and for display ads that succeed in attracting consumer attention are equal to one. In general, a lower CTR value would simply scale down the merchant's willingness to pay for sponsored positions and display ads under PPI bidding; meanwhile bidding on PPC is not affected by CTR in our setup. This modeling difference would not affect our analysis. PPC is slighly simpler to explain and more realistic.

The substitutability of search and display ads. Our framework assumes that both search and display advertising are purely informative. This implies that merchants view search and display ads as partial substitutes, which is critical to the organic search distortion results when publishers and ad intermediaries are symmetric. As discussed for compulsive consumers, the evidence in Blake et al. (2013) suggesting informative search advertising may not apply for display. More generally, there may be some persuasive brand advertising on the internet. If both display and search advertising were persuasive in a similar way, then even if advertisers wished to reach consumers multiple times, decreasing returns would generate substitutability. But in principle, some merchants may specialize into using just one advertising channel and they might even separate complementary branches of their advertising strategy between the search and display advertising channels.[64]

Empirical evidence is limited. An important, related study by Goldfarb and Tucker (2011) used a natural experiment based on "ambulance-chaser" laws restricting postal advertising by law firms to show that law firms do consider offline and online advertising to be substitutes. Hahn and Singer (2008) provide a survey that points to substitutability of the display and search advertising channels. The FTC (2007) took the controversial view that the markets are essentially independent, but the EU (2008) antitrust authority notes a trend towards substitutability for advertisers in large part as the technologies used for search and display advertising were and have continued to become increasingly similar (see point 52 of EU, 2008).[65] In particular, the EU (2008) found that some respondents considered

[63]Concretely, suppose each consumer only values one of the two merchants offering each product and can costlessly learn where a sponsored link leads, by observing a snippet. If we further characterize these two merchants by different probabilities of being the desired seller, this determines and equals that merchant's CTR. In the one slot setting, there would be no conflict of interest over this dimension, with both consumers and G favoring the highest CTR. Assuming G observes the CTR, G would weight the merchant bids by CTR since merchants with low CTR and high expected value per click would otherwise win too often in a PPC auction. Consumers would benefit. This would be similar to the modeling approach of Athey and Ellison (2011) who provide a number of interesting further results.

[64]If e.g., display ads complement search ads, G would wish to encourage display advertising with or without integration; this would push in the direction of removing organic search distortions in our baseline case, and induce distortions towards high α publishers in the extension of section 6.

[65]The legal debates surrounding Google reported there are multi-faceted and many extend beyond the scope of our

search and display ads to differ only in terms of the "triggering mechanism" which we call targeting. Our model demonstrates formally how targeting of display ads raises substitutability. Since display targeting technologies have been improving over time, we expect the substitutes assumption to become increasingly relevant. For instance, Ratliff and Rubinfeld (2010) question the FTC (2007) claims, citing market research studies that suggest that search and display advertising are being increasingly used for similar types of marketing, where initially display ads may have been used more often for building brand awareness compared to search ads which were arguably preferred for direct-response online sales.

Endogenous entry. Of course, search distortions depend heavily on the assumption that search engine competition is ineffective. We discuss endogenous entry of publishers and investments in quality content in the concluding discussion. Here we comment briefly on endogenous entry of merchants. If merchants face a cost of participating in online advertising, with a decreasing reverse hazard rate, just as for consumers, G would have to commit to leave merchants a rent. Otherwise no merchant would participate. In this two-sided market, consumer participation is increasing in merchant participation (and conversely), so G's incentive to distort organic search (which raises merchant rents) is likely to rise under separation. However, the qualitative effects remain very similar. With full monopolizing integration and full influence over display advertising, organic search distortions would disappear as before.

9 Concluding remarks

In this paper, we constructed a model that explicitly describes the workings of markets for both search and display advertising. The two modes of advertising are imperfect substitutes for merchants. So a monopoly search engine has incentives to distort organic search to make display advertising less effective, thereby increasing the value of sponsored search. At the same time, the search engine has incentives to distort sponsored search in favor of merchants with high willingness to pay, despite low consumer relevance. We characterized how these incentives interact and how they depend on market characteristics, such as the power of targeting technologies.

We also investigated varying types and degrees of integration. We showed how a monopoly search engine may monopolize the entire market by buying one intermediary in the display advertising market.[66] With symmetric publishers, such integration reduces the incentives to distort organic search, as well as sponsored search. Indeed, with full integration into publishing, incentives to distort organic results then fully disappear, but in general results are more nuanced.

paper; see van Loon (2012) for a legal discussion.

[66]Note that since buying DoubleClick in 2007 and AdMob in 2009 and Admeld in 2011, Google has overtaken Yahoo in the market for display advertising which was Yahoo's remaining strong point; see Learmonth (2011).

As discussed throughout the paper, G's relationship with publishers and display ad intermediaries has both vertical and horizontal features. Vertically, a more reliable search engine attracts more consumers to online search and channels them on to relevant publishers' websites, where consumer visits are an essential input in producing display ads by publishers and intermediaries. Unlike standard vertical relationships, the downstream firms (publishers) do not pay a price for this input, but, there is a clear analogy to how vertical integration typically reduces input prices, raising downstream firms' net benefits. Here, vertical integration raises publishers' net benefits, because the upstream search engine then increases the quantity and quality of its input of consumers to publishers. As usual, integration leads the engine to internalize the vertical externality, raising the quantity of input delivered to downstream firms, but here the engine does so by raising the reliability of its search service to consumers. As a result, the vertical integration effect raises consumer and total surplus.

Horizontally, both G and publishers offer competing advertising platforms for merchants. G's decisions induce a quality differentiation between publishers and G, by undermining the effectiveness of publishers' websites. By making its content search less reliable, G imposes a negative horizontal externality on publishers. This is also detrimental to consumers, so integration also benefits consumers from the horizontal perspective: integration of G into publishing or display advertising leads G to internalize this horizontal externality and raise the reliability of its organic search results.

However, when G integrates with or affiliates a non-trivial fraction of publishers, new motives to distort organic search results emerge. Partial integration reinforces G's incentives to steal business from non-affiliated publishers. More specifically, G distorts organic search when a consumer's best-match content is on a website owned by a non-affiliated publisher, diverting traffic to affiliated publishers. As with non-integration, the distortion increases G's revenues from sponsored search advertising by disrupting the effectiveness of display ads. Now, in addition, G raises the revenues of its affiliated publishers by diverting the distorted search traffic to them. If this new effect is relatively strong, this partial integration may result in lower consumer and total surplus.

These results might seem to advocate integration with full monopolization, but even neglecting pluralism and publisher incentives, discussed further below, our next result shows that such a conclusion would be premature. Full integration does avoid the problem of diverting traffic from non-affiliated to affiliated publishers, but publishers may be intrinsically heterogeneous. When publishers' websites vary in their effectiveness as ad platforms, internalizing externalities among publishers creates new motives for distortion. Under full integration, G has an incentive to divert traffic from less to more effective publishers to increase the aggregate value of display advertising.[67] Again, this new, negative effect may

[67]Under separation, G diverts traffic in the opposite direction, from more to less effective publishers. This suggests that integration with publishers will result in a more "commercial" or ad-intensive experience for consumers during content

dominate the positive effects of externality internalization so that integration results in lower consumer and total surplus. Partial integration with only the more ad-effective publishers compounds this with the negative welfare impact of the previous paragraph.

Ultimately, predicting which of these competing effects will dominate is an empirical question. In the discussion, we already explained how our model provides a structure for estimating the degree of substitutability between display and search advertising. Substitutability plays an important role in determining the risk of organic search bias. In particular, with non-integration, the risk of organic bias is low unless substitutability is substantial.

As noted in the introduction, we hope that the theory and model from this paper will provide a useful framework for further research. On the empirical side, defining and estimating bias or even just the factors conducive to bias, is a serious but exciting challenge. Our framework may be helpful in a number of ways. By demonstrating how the five principal sets of actors interact to determine the different types of search bias, the model indicates the range of variables relevant to an empirical estimation. In addition, our results point out where to look for some new types of bias, as well as how biases may interact. On the theory side, our framework could be extended in several directions to deal with alternative market structures, varying elasticity in merchant participation, on which we only touched briefly, and subtler technological assumptions regarding ad space, ad-effectiveness and ad targeting. Two extensions seem particularly straightforward.

Integration between the search engine and an ad intermediary or publishers may facilitate behavioral targeting where publishers use a consumers' past search queries to target display ads more effectively. In this case, integration would add the consequences of an increase in targeting precision, σ, to the integration effects described in sections 4 to 7. By analogy with proposition 4, the improved targeting raises incentives to distort both types of search, but the net welfare implication of this additional channel may be positive since targeting facilitates offline trade for consumers who can only conduct content searches.

In our model, the quality of publishers' content was exogenous. If we extend the model by letting consumers' utility from online content, u, depend on costly publisher investments, then the reduction in publisher revenues associated with integration of the monopoly search engine into ad intermediation would clearly lead to lower levels of investment with negative consequences for consumer and total surplus. Given the value of informative, entertaining, social and educational content sites on the web, explicit consideration of this new channel will provide an important step towards a more thorough welfare assessment of integration.

consumption.

10 APPENDIX

10.1 Proof of Proposition 5

The first-order conditions with respect to r^O and r^S can respectively be written as:

$$H\left(\overline{c}\right)\frac{(1-\eta)M\left(r^S\right)+\eta\delta m_1}{M\left(r^S\right)-\delta m_1}\left[\frac{u}{\sigma\alpha\left(1-\beta\right)}+\left(v_1-v_2\right)\left(1-r^S\right)\right]\gtreqless 1, \tag{18}$$

$$H\left(\overline{c}\right)\frac{v_1-v_2}{m_2-m_1}\left[(1-\eta)M\left(r^S\right)+\eta\delta m_1\right]\gtreqless 1, \tag{19}$$

where $\delta=0$ for the case with no integration and $\delta=1$ for the case with integration. Note that for both equations, the LHS is higher when $\delta=1$ than when $\delta=0$. Also when $\delta=1$, the LHS of (18) is higher than for (19) even if $u=0$.

Suppose that under no integration $r^S=r^O=1$. That requires that the LHS of both equations (18) and (19) are higher than 1 when evaluated at $\delta=0$ and $r^S=r^O=1$. So, they are also higher than 1 at $r^S=r^O=1$ when $\delta=1$. Consequently, $r^S=r^O=1$ is a candidate solution under integration. In fact, it is the only candidate. In any other alternative, either r^S or both r^S and r^O are lower than one, and this is inconsistent with the fact that the LHS of (19) is decreasing in both r^S and r^O. So the solution is unchanged.

Suppose that $0<r^S<1$ and $r^O=1$ under no integration. Then (19) holds with equality for $\delta=0$. If under integration, $r^O<1$ then $r^S=0$, since the LHS of (18) is higher than for (19). But again, this is inconsistent with the fact that the LHS of (19) decreases with r^S and r^O. So $r^O=1$ under integration. So r^O is unchanged, which, given that the LHS of (19) is higher with $\delta=1$ than with $\delta=0$, implies r^S must be higher than under no integration.

Suppose that $r^S=0$ and $0<r^O<1$ under no integration. If under integration $r^S>0$, then once again $r^O=1$. Instead if r^S remains at 0 under integration, then since the LHS of (18) under $\delta=1$ is higher than under $\delta=0$, r^O must be higher than under no integration.

Suppose that $r^S=0$ and $r^O=1$ under no integration. Similar to the last case, r^S either rises with r^O staying at 1, or r^S remains at 0, only now with r^O remaining at 1.

Finally, suppose that $r^S=r^O=0$ under no integration. This policy could remain optimal under integration, but not if the LHS of (18) evaluated at $\delta=1$ and $r^S=r^O=0$ is higher than 1, i.e., if

$$H\left(v_2\right)\frac{m_2}{m_2-m_1}\left[\frac{u}{\sigma\alpha\left(1-\beta\right)}+\left(v_1-v_2\right)\right]>1 \tag{20}$$

because r^O is then necessarily positive under integration. So (20) is a sufficient condition to rule out the possibility that $r^S=r^O=0$ remains an optimal policy. Since the LHS of (6) is higher than the LHS of (20), this proves that there exists a non-empty set of parameter values defined by conditions (6)

and (20), for which integration strictly improves the reliability of the search engine in terms of one or both types of search.

10.2 The size of γ and potential welfare losses from partial integration

Suppose that under no integration $r^S = r^O = 1$. Hence,

$$H\left(u+v_1\right)\left(1-\sigma\alpha\right)\min\left\{\frac{u}{\sigma\alpha\left(1-\beta\right)},\frac{v_1-v_2}{m_2-m_1}m_1\right\}\geq 1.$$

Under partial integration, we study the derivative of Π^G with respect to r^O_{NG} expressed in equation (12). Evaluating at $r^S = r^O_G = r^O_{NG} = 1$, the derivative will be negative (and r^O_{NG} distorted) if and only if:

$$\left(1-\gamma\right)H\left(u+v_1\right)\frac{u\left(1-\sigma\alpha\right)}{\sigma\alpha}<1. \tag{21}$$

We consider two extreme cases. First, suppose that $\sigma\alpha$ is very small. Then, condition (21) will hold only if γ is very close to 1. In this case, the loss associated with the distortion of traffic to non-affiliated publishers will be small since the fraction of non-affiliated publishers is also small. Next, suppose that $\frac{u}{\sigma\alpha(1-\beta)} < \frac{v_1-v_2}{m_2-m_1}m_1$, and moreover $H\left(u+v_1\right)\left(1-\sigma\alpha\right)\frac{u}{\sigma\alpha(1-\beta)}$ is very close to 1. In other words, under non-integration G is close to indifferent between setting $r^O = 1$ and a value slightly below 1. In this case, any value of $\gamma > \beta$ will satisfy (21), so that, unless $1-\beta$ is very small, the impact of traffic distortion can be substantial.

References

[1] Alaoui, L. and F. Germano (2013), "Time Scarcity and the Market for News." UPF mimeo.

[2] Argenton, C. and J. Prüfer (2012), "Search engine competition with network externalities." *Journal of Competition Law & Economics* 8(1), 73-105.

[3] Armstrong, M., J. Vickers, and J. Zhou (2009), "Prominence and Consumer Search." *Rand Journal of Economics*, 40(2), 209-233.

[4] Armstrong, M. and J. Zhou (2011), "Paying for Prominence." *Economic Journal*, 121(556), 368-95.

[5] Athey, S., and G. Ellison (2011), "Position auctions with consumer search." *The Quarterly Journal of Economics* 126(3), 1213-1270.

[6] Baye, M., and J. Morgan (2001), "Information Gatekeepers and the Competitiveness of Homogeneous Product Markets." *American Economic Review* 91, 454-474.

[7] Blake, T., C. Nosko and S. Tadelis (2013), "Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment." NBER Mimeo.

[8] Börgers, T., I. Cox, M. Pesendorfer, and V. Petricek (2007), "Equilibrium Bids in Sponsored Search Auctions: Theory and Evidence." mimeo.

[9] Brin, S. and L. Page (1998), "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Computer Networks and ISDN Systems* 30(1-7), 107-117.

[10] Broder, A. (2002), "A taxonomy of web search." *ACM SIGIR Forum* 36 (2).

[11] Chen, Y. and C. He (2011), "Paid Placement: Advertising and Search on the Internet." *The Economic Journal*, 121(556), F309–F328.

[12] de Cornière, A., and G. Taylor (2013), "Integration and Search Engine Bias." Mimeo.

[13] Edelman, B. and Z. Lai (2013), "Exclusive Preferential Placement as Search Diversion: Evidence from Flight Search." Harvard Business School NOM Unit Working Paper No. 13-087.
[14] Edelman, B., and B. Lockwood (2011), "Measuring Bias in 'Organic' Web Search." Mimeo.
[15] Edelman, B. and M. Ostrovsky (2007), "Strategic Bidder Behavior in Sponsored Search Auctions." *Decision Support Systems* 43, 192-198.
[16] Edelman, B., M. Ostrovsky, and M. Schwarz (2007), "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords." *American Economic Review* 97, 242-259.
[17] Edelman, B. and M. Schwarz (2010), "Optimal Auction Design and Equilibrium Selection in Sponsored Search Auctions." *American Economic Review* 100, 597-602.
[18] Eliaz, K., and R. Spiegler (2011), "A Simple Model of Search Engine Pricing." *The Economic Journal* 121, 329-339.
[19] Ellman, M., and F. Germano (2009) "What do the Papers Sell? A Model of Advertising and Media Bias." *The Economic Journal* 119, 680-704.
[20] Etro, F. (2012), "Advertising and Search Engines. A Model of Leadership in Search Advertising." *Research in Economics* 67 (1), 25-38.
[21] European Union, "EU (2008)," Article 8, Case No COMP/M.4731 – Google/DoubleClick. Regulation (EC) No. 139/2004 Merger Procedure.
[22] Federal Trade Commission, "FTC (2007)," Statement of Federal Trade Commission Concerning Google/DoubleClick, (FTC File No. 071-0170) 3 and 7.
[23] Goldfarb, A., and C. Tucker (2011), "Search Engine Advertising: Channel Substitution When Pricing Ads to Context." *Management Science* 57, 458-470.
[24] Greenspan, R. (2004), "Searching for balance." *ClickZ* (April 30), http://www.clickz.com/3348071.
[25] Greenstein, S. (2013), *Innovation from the Edges*, Princeton U. Press, forthcoming, Ch. 16.
[26] Hagiu, A., and B. Jullien (2011), "Why Do Intermediaries Divert Search?" *Rand Journal of Economics* 42, 337-362.
[27] Hahn, R., and H. Singer (2008), "An Antitrust Analysis of Google's Proposed Acquisition of DoubleClick." Working Paper.
[28] Hotchkiss, G., M. Garrison, S. Jensen (2005), "Search engine usage in North America." White paper, Enquiro, Kelowna, BC, Canada.
[29] Jansen, B. (2007), "The comparative effectiveness of sponsored and non-sponsored links for Web e-commerce queries." *ACM Trans.* Web 1(1) Article 3.
[30] Jansen, B., D. Booth, and A. Spink (2008), "Determining the Informational, Navigational, and Transactional Intent of Web Queries." *Information Processing and Management* 44 (3), 1251-1266.
[31] Jansen, B., and M. Resnick (2006), "An Examination of Searchers; Perceptions of Nonsponsored and Sponsored Links During E-commerce Web Searching." *Journal of the American Society for Information Science and Technology*, 1949-1961.
[32] Learmonth, A. (2011), "Google Just Blew Past Yahoo in Online Display Advertising; Amid a Fast-Growing Display Market, Search Giant Is Gaining Share." *AdAge Digital*, May 26.
[33] Lehaie, S. (2006),"An Analysis of Alternative Slot Auction Designs for Sponsored Search." EC06.
[34] Liu, D., J. Chen, and A. Whinston (2010), "Ex-Ante Information and the Design of Keyword Auctions." *Information Systems Research* 21, 133-153.
[35] Lorigo, L., M. Haridasan, H. Brynjarsdóttir, L. Xia, T. Joachims, G. Gay, L. Granka, F. Pellacini, and B. Pan (2008), "Eye Tracking and Online Search: Lessons Learned and Challenges Ahead." *Journal of the American Society for Information Science and Technology* 59 (7), 1041-1052.
[36] Ratliff, J., and D. Rubinfeld (2010), "Online Advertising: Defining Relevant Markets." *Journal of Competition Law and Economics* 6 (3), 653-686.
[37] State of Search, 2012, http://www.stateofsearch.com/wp-content/uploads/2010/12/Google-market-shares-around-the-world-infographic.jpg accessed Feb. 6, 2013.
[38] Tarantino, E. (2013), "A Simple Model of Vertical Search Engines Foreclosure." *Telecommunications Policy* 37(1), 1-12.

[39] Van Loon, S. (2012), "The Power of Google: First Mover Advantage or Abuse of a Dominant Position?" In Lopez-Tarruella A. (ed.) *Google and the Law: Empirical Approaches to Legal Aspects of Knowledge*, The Hague: T.M.C. Asher Press.

[40] Varian, H. R. (2007), "Position Auctions." *International Journal of Industrial Organization* 25, 1163–1178.

[41] Wang, J., and R. Day (2007), "The effects of attention inertia on advertisements on the WWW." *Computers in Human Behavior* 23(3), 1390–1407.

[42] White, A. (2013), "Search Engines: Left Side Quality versus Right Side Profits." *International Journal of Industrial Organization*, Forthcoming.

[43] White, A., and Jain, K. (2012), "The Attention Economy of Search and Web Advertisement." Mimeo, Tsinghua University.

[44] Wilbur, K. (2008), "A Two-Sided, Empirical Model of Television Advertising and Viewing Markets." *Marketing Science* 27, 356-378.

[45] Xu, L., J. Chen, and A. Whinston (2010), "Oligopolistic Pricing with Online Search." *Journal of Management Information Systems* 27(3), 111-141.

[46] Xu, L., J. Chen, and A. Whinston (2011), "Price Competition and Endogenous Valuation in Search Advertising." *Journal of Marketing Research* 48(3), 566-586.

[47] Yang, S. and A. Ghose (2010), "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?" *Marketing Science*, 30(1).