Online Display Advertising Markets: A Literature Review and Future Directions

Hana Choi, Carl F. Mela, Santiago R. Balseiro, Adam Leary

Abstract. This paper summarizes the display advertising literature, organizing the content by the agents in the display advertising ecosystem, and proposes new research directions. In doing so, we take an interdisciplinary view, drawing connections among diverse streams of theoretical and empirical research in information systems, marketing, economics, operations, and computer science. By providing an integrated view of the display advertising ecosystem, we hope to bring attention to the outstanding research opportunities in this economically consequential and rapidly growing market.

Keywords: display advertising • literature review • real-time buying • real-time bidding • programmatic buying

1. Introduction

According to the Interactive Advertising Bureau (IAB herein) report, display advertising (ad herein) totaled $49.8 billion in the United States in 2018, up from the $39.4 billion reported in fiscal year (FY) 2017. This double-digit growth rate (26%) is fueled by the upswing in mobile browsing, social media activities, video ad formats, and the developments in targeting technology. Digital display ad spending has surpassed even search ad spending and will continue its rapid ascent in share and significance.

Given the substantial economic importance of the display ad market, the objective of this paper is to summarize the existing literature from the perspective of each player involved in the display ad ecosystem. In doing so, we take an interdisciplinary view spanning diverse streams of theoretical and empirical research in information systems, marketing, economics, operations, and computer science. We focus on literature that explicitly emphasizes operations of the display ad market as opposed to a more general discussion on related topics, including the large body of literature on advertising, sponsored search ad, auctions, mechanism design, and online algorithms.

A handful of papers survey various aspects of online ad in general, covering search, display, classified, and other forms of ads. Ha (2008) reviews online ad research published specifically in six advertising journals. Evans (2008, 2009) discusses the evolution of Internet ad and provides industry perspectives, while Bucklin and Hoban (2017) focus on marketing models developed to improve decision making in Internet advertising. Due to the broad scope of these papers, the distinguishing characteristics of the display ad market (such as the issues arising from the coexistence of guaranteed and nonguaranteed selling channels) and its ecosystem are precluded.

Focusing on the display ad market, existing survey papers either discuss issues facing advertisers (demand side), publishers (supply side), or intermediaries. On the advertiser side, Tucker (2012) and Goldfarb (2014) review the economic literature on targeting and privacy concerns, and highlight their trade-off in ad effectiveness. On the publisher side, Korula et al. (2016) cover a broad range of publishers’ decisions in selling display ad, and review studies on allocating impressions and designing contracts/auctions. On the intermediary side, Muthukrishnan (2009) considers issues faced by ad exchanges, which are online platforms that match advertisers with publishers. One exception is Wang et al. (2017), who discuss recent algorithms and challenges in computational advertising from both advertisers’ (bidding strategies in real-time buying) and publishers’ (pricing and auction mechanism) perspectives, focusing on the nonguaranteed display ad selling channel.

Our survey differs in that we provide a comprehensive view of the display ad ecosystem, including
both guaranteed and nonguaranteed selling channels, and delineate the research progress gained from the eyes of all the players involved, including advertisers, publishers, and intermediaries. Furthermore, we take an interdisciplinary view and draw connections across disciplines. By providing an integrated view and bringing attention to the outstanding research opportunities, we hope to provide additional motivation for subsequent studies in this rapidly growing market.

The remainder of the paper is organized as follows. Section 2 overviews the ecosystem and highlights key unique features of the display ad market. Next, we organize existing literature by advertisers (Section 3), publishers (Section 4), and intermediaries (Section 5). In each subsection, we address key managerial decisions and discuss related research issues and opportunities. Section 6 characterizes some contextual drivers of display advertising, namely mobile and social media, and Section 7 concludes.

2. The Display Advertising Ecosystem

This section enumerates the players, the selling channels, and the key characteristics of the display ad market. Players’ objectives, roles, and available information are briefly discussed to place in context the research topics discussed in subsequent sections.

2.1. Players

The display ad market is a two-sided market: on one side, advertisers procure ad impressions on publishers’ websites to reach potential consumers; and on the other side, publishers with consumers’ impressions purvey ad inventory to advertisers with the highest valuations for those impressions. In between, intermediaries facilitate the match between advertisers and publishers by managing data and providing optimization tools and algorithms for serving ads.

2.2. Selling Channels

Ad exposures in display markets are sold via guaranteed and nonguaranteed selling channels. The guaranteed selling channel (so-called direct sale) involves the advance sale of a number of impressions at a fixed price. Nonguaranteed sales occur in real time via ad exchanges (so-called real-time buying (RTB)).

The nonguaranteed selling channel is estimated to be 35% (≈0.825 × 0.42) of the total display ad revenue in 2018.

The main characteristics distinguishing the guaranteed selling channel from RTB are (i) the pricing mechanism (fixed price in the guaranteed versus auction in RTB), (ii) the information available about the impressions (thus targeting ability), and (iii) the players involved (especially the level of intermediation). These distinguishing characteristics are described for each selling channel next.

2.2.1. Guaranteed Selling Channel.

The guaranteed selling channel involves the buying and selling of bundles of impressions through a guaranteed contract, prior to users’ arrivals on the publisher’s site. In the guaranteed selling channel, an advertiser and a publisher negotiate a fixed price pertaining to when, where, and how the ads will be displayed. These contractual arrangements guarantee the number of impressions to be delivered satisfying certain targeting criteria, at a negotiated fixed price (cost-per-mille (CPM)), during a specified time period (e.g., 1 million impressions to female users in SF during July for $5,000).

Advertisers often work with ad agencies to facilitate media planning/buying and the negotiation process. On the publisher side, there often is an in-house sales team that negotiates the terms and prices of the contract and manages the relationship with the advertisers. Many aspects of the contracting process are being automated with programmatic buying technologies. The so-called programmatic direct (or programmatic guaranteed, automated guaranteed) enables publishers to avail their guaranteed deals to advertisers with smaller budgets who otherwise would not meet the minimum ad spend required to allocate the limited resources of the sales team. The advent of programmatic direct thus decreases the role of sales teams and lowers advertisers’ search costs associated with finding prices from multiple publishers, which in turn would affect advertisers’ optimal ad buying strategies across selling channels (Section 3.2.1).

Because guaranteed contracts are drawn in advance, both advertisers and publishers rely on their expectations about future impressions and associated user characteristics (of those who will arrive during the planned campaign period) when making contractual decisions. Because targeting criteria are typically coarse, advertisers end up buying a bundle of impressions at a fixed price. For an advertiser, important questions are when/how should this guaranteed bundle of impressions be bought and which targeting criteria to use (Section 3.2). For a publisher, the questions are whether to sell impressions in advance as a bundle and how to price the guaranteed inventory (Section 4.1).

2.2.2. Nonguaranteed Selling Channel.

The nonguaranteed selling channel involves the buying and selling of single impressions in real time via programmatic buying. This channel is called RTB, because an advertiser’s buying decision is made immediately after the user arrives on the website. The ad inventory is nonguaranteed, because an auction takes place for the incoming impression and the advertiser’s ad is served to the end user only if the advertiser wins the auction. RTB enables publishers to monetize beyond the guaranteed contracted volume and
advertisers to buy impressions based on user-specific (e.g., cookie) information. Questions advertisers face in this market include how to calculate the value of an impression and bid, how to leverage users’ past behavioral data, and how to react to competitors’ bidding strategies (Section 3.2.3). Publisher questions include how to allocate an incoming impression across advertisers with guaranteed contracts and the highest bidder in RTB (Section 4.3).

The ecosystem in RTB is more complex than the guaranteed selling channel, with many specialized intermediaries lying in between. These intermediaries provide the fundamental infrastructure required for selling, buying, and serving ads in real time, typically within milliseconds. For example, ad exchanges, such as DoubleClick Ad Exchange, OpenX, and AppNexus Marketplace, provide centralized marketplaces by conducting an auction for each available impression in real time. Recognizing that programmatic buying is a difficult operational problem, demand-side platforms (DSPs), such as MediaMath, Rocket Fuel, and Google Marketing Platform: Display & Video 360, help manage placement and bids for the advertisers. Similarly on the publisher side, supply-side platforms (SSPs), such as DoubleClick for Publishers, Rubicon Project for Sellers, PubMatic, and MoPub, facilitate publishers’ inventory management and yield optimization. Although little research exists to date on the role of these intermediaries, the welfare of advertisers and publishers may improve by intermediaries’ enabling technologies that enhance targeting and allocation of impressions, although concerns exist regarding high commissions and incentive misalignment (Sections 5.1 and 5.2).

Finally, there are data intermediaries. Both advertisers and publishers strive to manage and leverage consumer data efficiently for targeting, but have access to different information. Advertisers, for example, can augment their own proprietary data (e.g., purchase patterns on the advertiser’s website) with additional data bought from third parties (e.g., income level, job history, home ownership, monthly car payment) to build targeting audience profiles. Publishers seek to integrate their own proprietary data (e.g., user registration information and ad viewing patterns on the publisher’s website) with third-party data to build user segments and to offer better targeting options to advertisers at potentially higher prices. To this end, data management platforms (DMPs), such as Adobe Audience Manager, Lotame, SAS Data Management, Oracle DMP, Nielsen DMP, and Salesforce DMP, provide an accessible interface to import data from multiple sources and build audience segments from the integrated data, which are then fed into DSP/SSP toward optimizing downstream decisions. For both advertisers and publishers, important research questions pertain to the optimal level of targeting granularity (to buy or to sell) and to which additional information to buy from third parties and at what price (Sections 3.3.3 and 4.2). Asymmetry of information also raises questions regarding how much information to share and the role of intermediaries in resolving this asymmetry (Section 5.4).

3. Advertisers

Media planning and execution constitute the demand side of the display ad market. Advertisers (i) set objectives (Section 3.1) and (ii) buy a set of impressions toward maximizing the objectives given their constraints (Section 3.2). Throughout the ad campaign, marketers (iii) measure return on investment (ROI herein) and use it as a reference point in setting future objectives and buying strategies (Section 3.3).

3.1. Objectives

Characterizations of display advertising have suggested two major goal types (e.g., Zhu and Wilbur 2011). One type involves building long-term brand equity. The other type is to generate short-term direct responses similar to coupons in print ads. As display ad objective is not directly observed by researchers in most data sets, integrated studies about what factors drive this twofold objective and the relative importance of goal types are scant. As display ad objective drives advertisers’ downstream decisions (such as budgeting and ad buying) and is antecedent to modeling, understanding this twofold objective is a pertinent area of future research.

There exists an array of key performance indicators (KPIs) used by advertisers to evaluate their campaign performance and to set (implementable) objective functions for optimization algorithms. The most common among these include impressions (Danaher et al. 2010), clicks (Chatterjee et al. 2003), visits (Dalessandro et al. 2015), and purchases (Manchanda et al. 2006). Consequent to the often incongruent twofold objective of increasing brand equity and enhancing direct response, the question arises as to which metrics should be evaluated and optimized. For example, should advertisers contract to buy a guaranteed number of impressions versus clicks? Does the time spent on a page and the number of page views constitute good measures for advertisers focusing on branding? Should survey-based metrics such as brand awareness, attitude, recognition, or recall be collected pre- and/or postcampaign (Goldfarb and Tucker 2011a, Bart et al. 2014)? If so, how can these data be leveraged in the optimization process? Are advertisers using the right metrics to drive profitability? Future research can further explore the relationship among advertisers’ display ad objectives, optimization processes, and various KPIs evaluated in practice.
3.2. Ad Buying

After the advertising objective is set, advertisers face the following ad buying decisions: (i) whether to buy guaranteed or nonguaranteed (Section 3.2.1), (ii) ad design and inventory characteristics (Section 3.2.2), (iii) user characteristics (targeting) (Section 3.2.3), and (iv) scheduling (frequency and timing) (Section 3.2.4). We discuss each decision in turn.

3.2.1. Selling Channels: Guaranteed Versus Nonguaranteed.

Most advertisers engage both selling channels by allocating a portion of their budget in the guaranteed selling channel in advance, and then spending the rest in RTB. Unfortunately, the existing literature on display ad is largely silent on why/when one selling channel is/should be chosen over the other. Next we discuss some plausible explanations, which future research can explore.

Some reasons why advertisers may favor RTB include finer targeting and lower search costs. When buying in real time, advertisers can submit an appropriate bid based on user-specific information at the impression level (e.g., user’s past behavioral pattern). Further, procuring a guaranteed contract entails time-consuming negotiation processes with potentially multiple publishers. These search costs are significantly reduced in RTB, as advertisers can reach a large number of publishers through a centralized intermediary exchange market.

On the other hand, guaranteed contracts are often favored when advertisers have a long-standing relationship with the publisher that facilitates the customization of ad formats and price negotiations (e.g., quantity discounts, bundling). If advertisers are risk averse, paying a premium to buy a guaranteed inventory in advance helps mitigate uncertainty in either the auctions’ outcomes or the amount of impressions that will be available in RTB on specific dates. Lastly, advertisers highly concerned with ensuring brand safety will choose guaranteed contracts so their ads appear on high-quality, reputable websites.

Athey et al. (2016) demonstrate that with consumers’ multihoming across publishers and imperfect tracking technology, advertisers seeking broader reach would favor larger publishers to avoid inefficient duplication, which can create an incentive for contracting guaranteed deals. As a supply-side explanation, Sayedi (2018) explores dynamic allocation (see Section 4.3), under which having both selling channels yields higher profits than selling in the guaranteed selling channel or RTB alone. The increase in profits arises because the presence of a guaranteed contract serves as a revenue alternative under dynamic allocation, especially when RTB auctions are thin.

A large number of positive and normative research questions remain regarding dual-selling channels. Positively speaking, an issue of interest is what the rationales are behind advertisers’ observed selling channel choices and their relative buying shares. For example, why would some advertisers invest more heavily in the guaranteed contracts and others more heavily in RTB? Why do dual-selling channels coexist in equilibrium? Normatively speaking, when would it be beneficial for the advertisers to choose one selling channel over the other, and what will be the optimal buying strategy?

3.2.2. Ad Design and Inventory Characteristics. In addition to the selling channels, advertisers make decisions regarding ad design and inventory characteristics. These involve choices about (i) websites (Danaher 2007, Danaher et al. 2010, Paulson et al. 2018, James et al. 2020) including traditional media (e.g., nytimes.com), blogs, and social media (e.g., Facebook); (ii) positions (Aksakalli 2012); (iii) sizes; (iv) creatives (Urban et al. 2013, Bruce et al. 2017); (v) formats including banner, text (e.g., Google AdSense), rich media (e.g., expanding creative), digital video, and native ads (Wojdynski and Evans 2016); and (vi) devices including desktop, mobile, and tablets.

Past research has typically evaluated each ad design and inventory characteristic independently to ascertain its respective effects on traditional brand equity measures (e.g., awareness, recall, recognition, attitude) and/or direct response measures (e.g., click-through rate purchase intent). However, relatively little effort has been apportioned toward assessing the interrelationship among these various inputs or their efficacy with respect to cost (Johnson and Lewis 2015). As ad design and inventory characteristics work jointly, optimization entails taking a combination of websites, positions, sizes, ad creatives, formats, and devices into account, and weighing the incremental benefits and costs of these options toward achieving the display ad goal given constraints (e.g., budget).

Furthermore, the choices of ad design and inventory characteristics are closely tied with the choices of selling channels. In RTB, standardization is necessary to ensure automation and seamless delivery within milliseconds. As a result, ad exchanges generally support inventories that conform to the IAB standard guideline, and customizability in terms of ad position, size, format, or content environment is limited. For example, highly customized ad positions and environments, such as homepage takeover (Sayedi et al. 2018) or conquest ads, are generally supported only via guaranteed contracts. Goldfarb and Tucker (2014) provide evidence that ad recognition declines as the use of IAB standard formats rises, presumably because standard format ads attract less attention and are harder to distinguish from competitors’ ads. This raises the question of how the level of standardization
or the demand for customizability affect ad buying decisions and market outcomes.

From an implementation perspective, scalability, computational efficiency, and accuracy are key considerations in building optimization algorithms. An optimization algorithm taking both selling channels into account is imperative. For computational efficiency, advertisers might be making these decisions hierarchically in practice (e.g., allocating budget across selling channels first, then across websites, then devices and formats, etc.) and an easy-to-implement optimization algorithm might be considered upon these heuristics.

### 3.2.3. Target User Characteristics

Advertisers often target ads to users whose purchase might be imminent (generating direct responses) or who might potentially become valuable customers in the future (building brand equity or customer lifetime value). Central to this goal is the identification of targets conditioned upon the available information. In demographic targeting, users are segmented based on age, gender, and geography. Contextual advertising, on the other hand, targets based on users’ interests, for example, by matching ad product category to the website content the user is browsing. In behavioral targeting, users are sought after based on their past behavioral patterns, for example, users who visited an advertiser’s website in the past. We begin with targeting in the guaranteed setting, then move to the discussion of targeting in RTB.

#### 3.2.3.1. Targeting Under Guaranteed Selling Channel

For behavioral targeting, user profiles are generated based on users’ topic interests (Ahmed et al. 2011, Trusov et al. 2016), conversion intents (Chung et al. 2010, Pandey et al. 2011, Aly et al. 2012), or brand affinity (Provost et al. 2009), and the preferred users are selected as the target audience. Contracting upon such preferred user segments several months in advance via guaranteed contracts, however, is often difficult as users’ characteristics (e.g., topic interests) vary considerably over time. From a publisher standpoint, committing to deliver a certain number of impressions of users demanded by the advertiser may increase the risk of under-delivery, as the publisher may not have enough information to make good predictions on those users’ future impressions. Although behavioral targeting based on the advertiser-side information (e.g., users identified as having high purchase intent based on previous visits to the advertiser’s website) is not common, some publishers offer behavioral targeting based on the publisher-side information (e.g., users with frequent visits to the publisher’s website). Presumably the publisher tries to price discriminate and increase profits by offering those user segments whose future impressions are well predicted by the publisher based on its own data, and whose exposures are highly valued by the advertisers.

In terms of demographic and contextual targeting, advertisers often have good insights on which demographic groups and contents are compatible with the ad product or service. For example, the advertiser may want to target users living close to their brick-and-mortar store locations (Johnson et al. 2016a). The question advertisers often face is whether it is profitable to pay for an additional layer of targeting. Although not much research has studied advertisers’ user targeting decisions in the guaranteed selling channel, this is an eminent area for future research, especially in conjunction with RTB. For example, when targeting in guaranteed contracts is limited, advertisers might opt to buy inventory in RTB auctions where targeting is finer. The shift in bargaining power between advertisers and publishers will affect how much information (targeting) is provided and the price and demand for such targeting in the guaranteed selling channel.

#### 3.2.3.2. Targeting Under Nonguaranteed Selling Channel

In RTB, advertisers have considerably more information and control over targeting. Each impression is sold separately in real time, meaning that the value of an impression can more readily be evaluated by matching the incoming user with the advertiser’s own data on users’ past behavior. In addition to the demographic and contextual targeting, advertisers greatly enjoy behavioral targeting in RTB.

Determining targeted optimal bids requires two components: (i) the advertiser’s own valuation for the targeted impression and (ii) distribution of the highest bids among competing advertisers. Aggregating targeted impressions into a campaign, the advertiser will calculate the optimal bids to maximize its overall utility (welfare) given the auction format.

We first review the literature that address issues in predicting advertisers’ valuations and the distribution of others’ bids in real time, then move to the discussion on the optimal bidding strategy in consideration with the advertiser’s display ad goal and constraints.

#### 3.2.3.2.1. Real-Time Prediction on Advertiser Valuations

Placing a bid entails predicting the advertiser’s value of an arriving impression in real time. Toward this end, prior research focuses on using immediate ad responses (mainly CTR, conversion rates) as proxies for valuation and builds algorithms to minimize prediction errors. Some computational challenges in predicting direct ad responses in real time include that (i) the dimensionality of user attributes space is
high and data are sparse across these dimensions, (ii) conversion events occur rarely, and (iii) cold-start situations with limited data history hinder reliable predictions. In general, the literature has evolved to address these challenges and to lower computational costs while improving predictions. Linear models (Chen et al. 2009; Agarwal et al. 2010, 2014; McMahan et al. 2013; Chapelle et al. 2015) are widely used in practice as they are easy to implement and perform effectively even with large-scale data sets. As it is difficult to capture higher order interactions with linear models, hybrid methods are being developed, incorporating factorization (Menon et al. 2011, Oentaryo et al. 2014, Juan et al. 2016) or decision trees (He et al. 2014), and deep learning methods are being applied to further explore latent patterns (Zhang et al. 2016b). The learnt user characteristics from user profiling (Perlich et al. 2014, Zhang et al. 2014) and features additionally inferred from social data (Bagherjeiran and Parekh 2008, Liu and Tang 2011, Goel and Goldstein 2013) are also leveraged in improving ad response predictions.

An inherent presumption in much of the preceding literature is that valuations are proxied by immediate conversion probabilities. However, this practice suggests serving ads to consumers who are likely to convert even without ad exposures. A better bidding strategy would, for example, instead use incremental expected payments. A bidder’s winning probability also determines the total expected number of impressions won and the reach metric for a given campaign. Challenges in predicting competing advertisers’ highest bids include that (i) impressions possess demographic, contextual, and behavioral attributes for which competing advertisers have heterogeneous valuations, and (ii) because competitors’ bids are not disclosed, the highest of the competing bids (and the reservation price set by the publisher) is only observed from the winning payment, and the lower bound of competing advertisers’ highest bids is inferred upon losing from the advertiser’s own bid.

3.2.3.2.3. Optimal Bidding Strategy. Advertisers seek to determine their optimal bids given the auction mechanism faced. Until recently, ad exchanges typically conducted second-price, sealed-bid auction for each impression available, and much of the existing display advertising literature thus far studies this auction format (or a variant thereof). In theory, the weakly dominant strategy for an advertiser in a private value, second-price auction of a single object is bidding truthfully. However, bid calculation becomes more complex as additional practical constraints are considered: the advertiser (i) faces a budget constraint for a given campaign in repeated auctions (Balseiro et al. 2015, Balseiro and Gur 2019), (ii) learns own and/or other’s true valuations over time (Iyer et al. 2014, Cai et al. 2017), (iii) sets a number of impressions to attain (Ghosh et al. 2009b, Choi and Mela 2018), and (iv) sets pacing options so that the budget is spent smoothly over a specified time period (Lee et al. 2013, Yuan et al. 2013, Xu et al. 2015). With such considerations, the optimal bidding strategy is not necessarily truth telling. For example, Balseiro et al. (2015) show that the optimal bidding strategy for an advertiser facing a (binding) budget constraint is to shade values to account for the option value of future opportunities.

In designing optimal bidding algorithms, different assumptions have been made about the advertiser’s objective function and the bidding environment, including the ways predictions are made about valuations and the distribution of highest bids. One common approach predicts statistics of interest (e.g., CTR for ad responses, moments for the distribution of highest competing bids) as a preliminary step toward minimizing prediction errors, and then using these predictions as inputs in calculating the bids (Perlich et al. 2012, Zhang et al. 2014, Xu et al. 2016). However, when advertisers are highly uncertain about ad responses or the distribution of the highest bid, incorporating the prediction step into the entire bidding optimization task is desirable to enhance learning about the overall utility from a campaign (Chapelle 2015, Ren et al. 2016, Hummel and McAfee 2017).

Because training the model is computationally expensive, training has traditionally been done offline to find the optimal bidding strategy, which is then applied to the real-time data. As the campaign progresses, advertisers receive additional data on winning outcomes with corresponding payments.
and user responses. To incorporate such new information or to dynamically adapt to the changes in the competition environment, while moderating computational cost, research considers shortening the successive offline optimization cycles by applying simple algorithms that can myopically reoptimize (Lang et al. 2012) or by tuning the bidding parameters dynamically according to the current performance feedback (Cai et al. 2017).

Developing an adaptive and better performing real-time bidding algorithm is an ongoing topic of interests (e.g., Tunuguntla and Hoban 2018). In this regard, the participation and transaction cost incurred in RTB are yet to be explored. At the selling channel level, advertisers face entry costs in terms of understanding the complexity in RTB and setting up an appropriate bidding process. At the auction level, there are transaction costs such as monetary fees charged by the intermediaries and/or cognitive effort costs in bidding. Although we are aware of no research quantifying the participation and transaction costs in RTB, these costs will considerably affect advertisers’ participation decisions, competitive landscape, and resulting auction outcomes.

3.2.4. Scheduling. We next consider advertisers’ ad scheduling decisions, that is, when to display ads and how often. Finding optimal frequency is important not only to attain the highest ad effectiveness (Johnson et al. 2016a) but also to optimally allocate limited resources. Timing is important due to the long-term carry-over or spacing effect of ads (Braun and Moe 2013, Sahni 2015, Sahni et al. 2019).

In the guaranteed selling channel, publishers mostly control the ad delivery schedule and commonly offered decision variables are frequency capping and pacing options. An implementable algorithm that can choose these common decision variables (e.g., how to set frequency capping, does uniform pacing result in better exposure timing) toward achieving optimal scheduling will be valuable.

In RTB, where each impression can be bought separately based on the user-specific information, advertisers have better control over frequency and timing at the individual level, conditioned upon users’ browsing behaviors. To this end, one possible direction for future research is building an optimal bidding algorithm that takes the effect of frequency and timing into account (e.g., carry-over effect of ad) in evaluating own valuations of an impression and calculating the bid. Having a greater control over individual-level frequency and timing might be an additional rational for advertisers to allocate budget toward RTB. As winning is probabilistic in RTB and the bid landscape changes rapidly, ad scheduling is an interesting dynamic control problem.

3.3. Measuring Value of Ad Spend

Next we discuss challenges in measuring the actual posthoc value of ad spend to the advertiser (i.e., causal effect as opposed to the advertiser’s ex ante prediction for the ad valuation) and suggest some directions for future research.

3.3.1. Attribution. Assessing and quantifying the effectiveness of each ad is remarkably difficult, as the consumer is exposed to an advertiser multiple times via various channels, and as these “multi-touches” jointly influence consumer behavior. For example, display and TV ads simultaneously affect online and offline sales. How advertisers attribute the increase in desired outcome to each ad exposure affects how advertisers would update budget allocation across media channels and also the buying decisions in the display ad market (Jordan et al. 2011, Shao and Li 2011, Abhishek et al. 2012, Dalessandro et al. 2012, Li and Kannan 2014, Xu et al. 2014, Kireyev et al. 2016, Berman 2018). With the advancement in tracking technology, richer datasets become available that connect ad exposures not only across communication channels (display, search, TV) but also across devices (desktop, mobile, tablets). Individual-level online activities are further being matched to offline behaviors, for example, via reward cards. Ongoing progress is being made to capture partial aspects of these cross-channel effects (e.g., effect of display ad on online and offline sales in Lewis and Reiley 2014; effect of search and display ads on online and offline purchases in Abraham 2008; effect of display ads on consumer searches in Papadimitriou et al. 2011, Lewis and Nguyen 2015, Ghose and Todri-Adamopoulos 2016; effect of mobile and desktop exposures on mobile and desktop conversions in Ghose et al. 2013) and future research is warranted encompassing the entire customer journey (i.e., effects of multichannel exposures on multichannel behaviors) to better understand attribution and the resulting implications on ad buying strategies.

3.3.2. Econometric Issues and Experimental Design. Putting aside the attribution and cross-channel issues, measuring the causal effect in general poses several challenges, including selection biases arising from targeting and browsing intensity (see Lewis et al. 2015 and references therein, Hoban and Arora 2018). Observational variation, even with the most sophisticated estimation methods, can easily fail to yield reliable measurement of ROI (Gordon et al. 2019). To overcome the econometric issues and to better understand the value of ad spend, both advertisers and researchers are turning attention to randomized field experiments. The key question is how to design experiments to obtain unbiased estimates with
statistical power while reducing experimental cost (Kohavi et al. 2009, Barajas et al. 2016, Johnson et al. 2016b, 2017a, 2017b). Design feature options include the randomization method (e.g., randomize before targeted users are selected or after, sample sizes for the control and treatment groups, type of ads to be shown to the control group (e.g., placebo ad or competing advertiser’s ad), length of the experiment(s), and sample data period (i.e., data collection period prior and/or post experiment)).

Using the results from the experiments, advertisers ultimately want to infer the relative effectiveness of different ad buying strategies and optimally reallocate resources. For example, Hoban and Bucklin (2015) calculate marginal effects and elasticities along the purchase funnel, and suggest ways to reallocate impressions across users at different stages. Similarly, Barajas et al. (2016) compare performance-based campaign with CPM-based campaign and leverage the results in reselecting target users. Presumably advertisers seek to test different ad designs and inventory characteristics, targeting types, targeting algorithms, frequency and timing of ad exposures, and so forth. Based on the result from the previous experiment, how should the next experiment be designed? Related questions are in which order these decision variables need to be tested and which variables should be grouped together in the testing environment to reduce the experimental cost. There exists an increasing demand for automated A/B testing solutions, and more work would be welcome in the context of designing and effectively learning through repeated testings toward finding the global optimal ad buying decision. Repeated feedback cycles (i.e., testing → analyzing results → updating → retesting) become more crucial when the display ad environment changes rapidly over time and when competitors’ ad strategies change in response to focal advertiser’s reoptimization.

3.3.3. Advertising Effect and Privacy. More precise targeting generally increases ad effectiveness by delivering ads that are in consumers’ interests (Beales 2010, Goldfarb and Tucker 2011b, Farahat and Bailey 2012, Lambrecht and Tucker 2013, Bleier and Eisenbeiss 2015). However, finer targeting requires some degree of privacy intrusion, and may backfire as customers’ sense of vulnerability and privacy concerns increase (Goldfarb and Tucker 2011a, Tucker 2014b, Aguirre et al. 2015). More stringent privacy policies (opt-in < opt-out < tracking ban) negatively impact profitability of advertisers and publishers, and there is still an ongoing debate over regulatory controls (Johnson 2013, Johnson et al. 2020). How privacy policies should be set in consideration with consumers’ welfare (benefits from increased protection of privacy, minus loss from being exposed to less relevant ads) and changes in welfares of advertisers, publishers, and intermediaries is an interesting area for future research. Another important area for future research concerns how much data to use in targeting and the value of different types of data bought from third parties, when considering the trade-off between finer targeting and privacy concerns. For example, it is unclear whether and when advertisers benefit from acquiring information on users’ social graphs and leveraging it in behavioral targeting.

4. Publishers
A publisher on the supply side of the display ad ecosystem faces the following decisions: (i) pricing advertising inventory (Section 4.1), (ii) deciding what information to share and with whom (Section 4.2), and (iii) scheduling the ad delivery (Section 4.3).

4.1. Setting Prices
Advertisers jointly set prices conditioned on advertiser valuations across the guaranteed selling channel and the RTB channel. However, most prior literature explores publisher pricing decisions conditional on the selling channel, and we address some challenges inherent in the joint pricing problem.

4.1.1. Inferring Advertiser Valuations. It is self-evident that a first step in setting prices across channels is understanding advertiser valuations as discussed in Sections 3.2.3 and 3.3. That said, publishers face additional challenges in ascertaining their respective advertisers’ valuations as publishers do not observe advertisers’ proprietary data (e.g., purchase conversion rate). Eliciting advertisers’ valuations based on the limited information available to publishers, and how these may vary across selling channels, is an important area for future research. For example, joint consideration of the bid amount in the exchange auction and the fixed price paid in the guaranteed contracts and the volume bought in each selling channel can provide more information about advertisers’ valuations.

4.1.2. Pricing in Guaranteed Selling Channel. In the guaranteed selling channel, publishers and advertisers come to an agreement on a fixed price pertaining to when, where, and how the ads will be displayed. We first discuss the most common CPM pricing, then explore other forms of pricing mechanisms.

4.1.2.1. CPM Pricing. The difficulty in committing to guaranteed contracts in advance is that publishers face uncertain demand (advertisers’ arrivals) and ad inventory (viewers’ arrivals) (Feige et al. 2008). To address these uncertainties, existing research typically
constructs a revenue management model in which advertisers arrive sequentially with desired quantities of impressions and valuations, and in which the publisher makes contractual decisions prior to knowing other demands arriving in the future. The publisher either (i) determines whether to accept arriving advertiser’s proposed price and quantity, where the accepted request may be canceled at a cost (Babaioff et al. 2009, Constantin et al. 2009, Roels and Frideirsdottir 2009), or (ii) dynamically determines a price to quote to an arriving advertiser’s demand (Frideirsdottir and Najafi-Asadollahi 2018).

The actual guaranteed contractual process more closely resembles a bargaining process, where the outcome may depend on several factors including advertiser’s willingness to pay, publisher’s outside option, competition, business relationship, and negotiation skills. All these factors suggest interesting questions for research. As an example, little is known about the role salespeople play in selling guaranteed contracts. Salespeople might lower the search costs of advertisers in finding the right website, or convey additional information and value about the impressions to the advertisers. With the increase in programmatic direct, the role and value of salespeople may change.

Finally, as advertisers have different valuations for varying inventory characteristics (e.g., size, format), publishers can bundle some of these ad units (e.g., desktop+mobile) to better price discriminate. Which ad units to bundle at which fixed bundle price is an open question.

### 4.1.2.2. Other Pricing Schemes

Although CPM pricing is predominant in display ad, publishers can also consider other pricing schemes such as CPC (Mangani 2004, Fjell 2009, Asdemir et al. 2012, Najafi-Asadollahi and Frideirsdottir 2014) or subscription fees (Baye and Morgan 2000, Kumar and Sethi 2009). Kumar and Sethi (2009), for example, develop a dynamic hybrid revenue model that determines the optimal subscription fee level and amount of ad on a website over time, in consideration with the cost of serving each customer, cost of presenting ads, and cost of changing the subscription fee. Another interesting approach is taken in Goldstein et al. (2015), in which a time-based pricing scheme is considered because duration of an ad affects memory retention. Future research is warranted in exploring and comparing different pricing mechanisms that further look at the effect of advertisers’ competition, publishers’ competition, and consumers’ multihoming behaviors across websites.

### 4.1.3. Pricing in Nonguaranteed Selling Channel

Compared with the large body of literature focusing on advertisers’ optimal bidding strategies, research on publishers’ optimal pricing policies in RTB are relatively scarce. As in guaranteed contracts, information asymmetry (Hu et al. 2016) or double marginalization (Dellarocas 2012) can affect the publisher’s choice of pricing type in RTB, that is, CPM, CPC, or other cost-per-action (CPA) auctions. Existing literature further considers publisher’s incentives in adopting a hybrid format where advertisers are allowed to choose between CPM or CPC type auctions (Zhu and Wilbur 2011, Liu and Viswanathan 2013).

Although intermediaries typically control many aspects of the auction mechanism (e.g., second-price versus first-price auction format, see Section 5.3), the publisher still faces its pricing decision conditioned on the auction rules. For example, adding reserve prices in RTB auctions can substantially increase publishers’ revenue, especially when asymmetric, heterogeneous advertisers compete in thin auctions (Yuan et al. 2014, Paes Lemke et al. 2016). The optimal reserve prices would need to take into account that advertisers’ buying strategies reflect some practical constraints they face such as budget (Balseiro et al. 2015) or reach (Choi and Mela 2018). When the distribution of bids is not known by the publisher, Amin et al. (2013), Medina and Mohri (2014), and Cesabianchi et al. (2015) explore how to learn optimal reserve prices in a nonparametric fashion. An open issue is setting the optimal reserve prices in the face of competing publishers whose inventories are considered as (imperfect) substitutes by advertisers, and to develop algorithms to deploy the reserve prices at scale into ad exchange auctions.

Private marketplaces in the nonguaranteed selling channel are becoming increasingly popular and different from open auctions inasmuch as only select advertisers can participate. Private marketplaces typically consist of prenegotiated, yet nonguaranteed, contracts for impressions. A common format is to provide priority advertisers access to invitation-only auctions inventory at a fixed CPM prior to impression being sold in the open auction. These ad contracts (often called preferred deals or first look) are analogous to option contracts in financial markets (Chen and Wang 2015) because the advertiser has the right, but not the obligation, to buy ad impressions at a fixed rate in the future. Despite the added complexity of this format relative to open auctions, advertisers are asking for private deals to reduce uncertainty in payment prices, brand safety concerns, to ensure a wider reach of impressions, and to flexibly access a coveted premium audience data (without having to commit to upfront contracts). Publishers often provide valuable audience segment targeting based on their own proprietary data for preferred deals, but not for open auctions. Different levels of information revealed can be a powerful product differentiator and
publishers can improve revenue with these deals (Mirrokni and Nazerzadeh 2017). However, a long list of questions remains regarding this nascent and growing pricing format, including when/how to offer these deals and at what prices, to whom to give private auction access, and how to set reserve prices in private auctions when the unsold impression can subsequently be sold at an open auction.

4.1.4. Interrelation of Guaranteed and Nonguaranteed Selling Channels. As advertisers can enter in both selling channels and publishers can sell in both, pricing should ideally be optimized jointly (Chen et al. 2014). When publishers are considering the interrelated benefits and costs of selling in both channels, it remains unclear what the jointly optimal guaranteed fixed prices and reservation prices should be. Moreover, other pricing mechanisms might yield even higher joint profits for publishers in this dual-selling channel structure by further price discriminating based on advertisers’ valuations or uncertainty. For example, a publisher can consider selling call-options that can be transferred among advertisers, who may pay a premium to hedge against fluctuating prices in real-time auctions.

4.2. Information

What information to share with the advertisers and at what price are important questions for the publishers and warrant further studies. In the guaranteed selling channel, the publisher faces a trade-off in providing more information to the advertisers and making more granular level targeting options available. On the one hand, the publisher can facilitate price discrimination as advertisers are willing to pay a premium for an additional layer of targeting. On the other hand, the publisher can choose to disclose less information by bundling impressions together. When the publisher sells a large number of ad impressions and does not know advertisers’ heterogeneous valuations for these impressions, predicting advertisers’ valuations for a bundle of goods is easier because, from the law of large numbers, the distribution of valuations for the bundle tends to concentrate around its mean (see, e.g., Bakos and Brynjolfsson 1999). Better predictions can, in turn, lead to higher revenues for the publisher. Additionally, coarser targeting criteria can reduce the risk of not meeting the contractual terms and paying a penalty in the event that the publisher falls short of the impressions. Thus, the publisher accounts for the benefits and costs of providing more information to the advertiser when underwriting the targeting terms in guaranteed contracts.

Similarly in RTB, the publisher may benefit from providing more information as advertisers bid higher amounts conditional on participating in the auctions (De Corniere and De Nijs 2016, Hummel and McAfee 2016, Ada et al. 2019). At the same time, providing more information decreases the number of participating bidders and creates thin markets, as fewer advertisers are interested in a given impression with highly differentiated attributes (Levin and Milgrom 2010, Fu et al. 2012, Chen and Stallar 2014). In the presence of advertisers’ and consumers’ heterogeneous preference for targeting, Gal-Or et al. (2018) show that the differentiation in targeting levels offered to advertisers in equilibrium depends on the intensity of competition among platforms who face multihoming consumers.

Providing more information, thus finer targeting ability, also results in displaying ads that highly coincide with users’ recent behaviors, which may increase privacy concerns and reduce ad responses (Lu and Yang 2018). Furthermore, advertisers can use shared information to target the same user on other, cheaper websites (Ghosh et al. 2015). Because of these trade-offs, a publisher’s decision to share (Emek et al. 2014) or to price and sell the publisher’s own proprietary data are more complex than that of a third party selling cookies (Bergemann and Bonatti 2015).

This raises questions regarding whether publishers should share information about users’ behaviors prior to and/or after the ad exposures, and whether it will benefit publishers to share KPIs for each specific type of ad inventory or as overall averages.

4.3. Delivery of Ad and Impressions

Ad scheduling is an inherently challenging task for the publisher. On one hand, publishers have guaranteed contracts to fulfill. On another hand, publishers can potentially enjoy high bids from those advertisers who want to cherry pick impressions in RTB. In addition, there is an uncertainty about the supply of impressions, which are largely determined by users’ browsing behaviors.

For the guaranteed selling channel, prior research has considered allocating arriving impressions among contracted advertisers, seeking to maximize publisher revenue (Adler et al. 2002, Kumar et al. 2006, Feldman et al. 2009, Vee et al. 2010, Devanur et al. 2011, Ciocan and Farias 2012, Turner 2012) and/or ad responses (Chickering and Heckerman 2003, Korula et al. 2013) when contracts differ in prices, campaign schedules, targeting criteria, and delivery status. As RTB markets have grown, there has been increased interest in dynamic allocation, where publishers first request bids for each impression (often by adopting header bidding) and compare the winning price to the option value of assigning the impression to the best matching guaranteed contract (Ghosh et al. 2009a, Balseiro et al. 2014, Arnosti et al. 2016, Chen 2017, Sayedi 2018). These studies show that dynamic allocation can yield
higher profits for the publishers than first sending (random quality) impressions to the guaranteed contracts. Nevertheless, many publishers still prefer selling premium high-quality inventories (e.g., front page, leaderboard) via guaranteed contracts, and limit the application of dynamic allocation to lower quality inventories with lower guaranteed contract prices. Presumably for certain types of inventories, leveraging advertisers’ willingness to pay high premium for reach or brand safety (ensured via guaranteed contracts) benefit publishers more than leveraging high bids for cherry-picked impressions in RTB. In sum, allocation algorithms remain an active area for inquiry and are likely incumbent upon how advertisers value upfront contracts.

In addition, publishers could allow budget pacing (Balseiro et al. 2017a, Conitzer et al. 2019), impression pacing (Bhalgat et al. 2012), or frequency capping (Hojat et al. 2017), and assess how these options provided to advertisers affect publisher’s optimal ad delivery plan. Another question relates to controlling supply of impressions to facilitate ad delivery or to raise advertiser competition. Publishers could consider buying additional traffic (e.g., via Facebook ad) to intentionally increase supply, or intentionally not sell certain ad spaces at a given point in time. Varying (or smoothing) supply of impressions affects advertiser competition and allocation of impressions among advertisers, making this another potential direction for future research.

5. Intermediaries

As outlined in the display ad market ecosystem discussion, many intermediaries (DSPs, SSPs, ad exchanges, ad networks, data aggregators) serve technological needs in RTB and facilitate the match between advertisers and publishers. These intermediaries affect market outcomes, yet research remains sparse regarding the implications of intermediaries’ decisions. We discuss (i) welfare implications of enabling technologies (Section 5.1), (ii) principal-agent incentive misalignments (Section 5.2), (iii) intermediaries’ decisions regarding pricing (Section 5.3), and (iv) whether competing intermediaries can benefit from sharing information such as bids and payments (Section 5.4).

5.1. Enabling Technologies and Vertical Integration

The evolution of buying and selling practices in display advertising are tightly integrated with the development of intermediaries’ enabling technologies (e.g., data collection, data analysis, RTB, real-time ad serving, A/B testing and optimization tools). With such technologies, intermediaries can better match publishers’ impressions (consumers) to more relevant advertisers (Broder et al. 2007, Chakrabarti et al. 2008, Zhang and Katona 2012). Further, Balseiro and Candogan (2017) show that intermediation leads to an increase in the overall market efficiency when optimal contracts are offered to budget-constrained advertisers with private information on budgets and targeting criteria.

On the other hand, these intermediaries charge high fees for their services. Although the fees charged by these ad tech intermediaries vary widely depending on the level of services provided in buying and selling ads, their revenues constitute approximately 55% of the ad spend in RTB. High intermediation fees have led a large fraction of brand advertisers to execute programmatic buying in-house. However, high costs associated with sourcing highly skilled media procurement talent to manage the time and effort required in buying ads has led some advertisers to turn back to their agencies for help. Overall, the welfare implications of intermediaries who enable new practices in display ad markets (e.g., existence of RTB and targeted ad buying at scale) should be examined.

5.2. Agency

The existence of intermediaries (with their enabling technologies) raises interesting research questions as economic incentives of principals and intermediaries may not be aligned. Balseiro et al. (2020) explore how the structure of an intermediation network affects the profits of participants when advertisers’ values are private, and show that intermediaries have incentives to shade bids and not to allocate impressions, even when profitable for their downstream customers. Allouah and Besbes (2017), on the other hand, explore the welfare implications of the collusive behavior of DSPs who represent multiple advertisers, and show that moving from collusion (DSP submitting a single bid to limit competition among the advertisers it represents on a given impression) to noncollusion (DSP bidding for each advertiser independently of other advertisers it represents) leads to a Pareto improvement when taking into account the publisher’s reaction such as increasing reserve prices. Future research is called for looking at different objectives of intermediaries, which might differ inherently from those of the principals, and appropriate contract mechanism design.

5.3. Pricing Mechanisms

In this subsection, we first discuss existing considerations regarding intermediaries’ pricing (i.e., auction mechanism design) on revenues of advertisers and publishers. We then outline new developments and pricing challenges (such as those associated with dynamic auction design and header bidding).

One concern spanning the existing literature is that ad impressions are highly differentiated, valuations are heterogeneous across advertisers, and the
distribution of advertisers’ valuations often violates the regularity conditions in Myerson (1981).\(^{33}\) Moreover, with user-specific information available in RTB, it is likely that advertisers have private information suggesting that certain types of impressions are commonly valued (Abraham et al. 2016). Ad exchanges further face asymmetric information states regarding publishers’ opportunity cost of selling the impressions and competition (Gomes and Mirrokni 2014). With such characteristics, the revenue equivalence theorem fails to hold, and ad exchanges can improve their revenues by designing mechanisms that are different from the commonly used second-price, sealed-bid auction. For example, Celis et al. (2014) proposes a randomized sales mechanism that outperforms an optimal second-price auction, where the buyer can choose to buy-it-now at a fixed price or take-a-chance in an auction with random allocation among top \(d > 1\) bidders.

Going forward, there has been an increase in interest regarding the design of dynamic auctions. Because most advertisers participate in multiple auctions throughout their campaign, the auctioneer can offer dynamic mechanisms that link incentives across auctions and improve revenues compared with repeated static auctions (Balseiro et al. 2017b, Mirrokni et al. 2018). Further, the commission structure among intermediaries is likely to affect the intermediaries’ incentives, the auction formats they adopt, and the associated market outcomes. More attention is warranted to ascertain this dependency.

Competition among advertisers and publishers with multiple (competing) intermediaries’ pose additional complexity in display markets. One recently introduced coordination mechanism is header bidding, which gathers bids from multiple ad exchanges.\(^{34}\) Although different sets of advertisers (DSPs) interact with different ad exchanges, header bidding enables publishers to coalesce demand for the inventory and increase revenue by intensifying competition among ad exchanges and advertisers.

Header bidding can be construed as a two-stage sequential game: in the second stage, ad exchanges independently run second-price auctions; and in the first stage, the publisher chooses to deliver the impression to the ad exchange paying the highest price. The outcome from this game can drastically differ from that from a single second-price auction with all advertisers’ bids on one intermediary platform. For example, the highest bidder is not guaranteed to win the impression because most intermediaries submit their second-highest bids (= payments under second-price auctions).\(^{35}\) To ensure that the highest bidder wins the impression under header bidding and to increase transparency in payments, many ad exchanges are now instead adopting first-price auctions (or soft floors, Zeithammer 2019).\(^{36, 37}\) In spite of the rapid growth in header bidding, research is virtually nonexistent regarding how advertisers compete in a multistage auction game with intermediaries competing toward increasing their own commissions (on payments by advertisers), and the attendant impact on the welfare of advertisers, publishers, and intermediaries.

Recently, Despotakis et al. (2019) provide a theoretical rationale to explain ad exchanges’ moves from second- to first-price auctions and the attendant revenue consequences for publishers and ad exchanges.\(^{38}\) They show that each ad exchange has a unilateral incentive to move to first-price auctions, given that publishers transition from traditional waterfalloing to header bidding.\(^{39}\) Under header bidding, first-price auctions lead advertisers to shade bids less aggressively and exchanges’ revenues to increase compared with second-price auctions. This follows because in first-price auctions the number of simultaneously competing advertisers increases. An unexpected consequence of this switch is that ad exchanges become now indistinguishable from the perspective of advertisers, because they can access any ad inventory from any exchange (with header bidding) and face the same auction outcome (as all ad exchanges run first-price auctions). The authors conjecture that, as a result, ad exchanges’ equilibrium fees to advertisers should decrease to zero once advertisers’ choices of ad exchanges are endogenized. In contrast, publisher revenues increase as competition intensifies among ad exchanges and advertisers. Empirically, it remains an open question to measure the magnitude of these welfare impacts on the participants in the ad ecosystem, especially in light of industry practices including advertisers’ constraints in bidding (e.g., budget, reach), complex fee structures, proprietary technological and contractual ties among ad exchanges, DSPs, and SSPs, and the ad exchanges’ ability to differentiate by leveraging information on ad impressions, bid landscape, and auction outcomes.

5.4. Information

Section 4.2 discusses the publisher’s problem with regard to information, and an analogous problem exists for intermediaries. Intermediaries possess information spanning advertisers and/or publishers and can therefore assess the competitive landscape more completely than a single advertiser or publisher. DSPs have access to ad budgets, bidding strategies, winning bids, and payments across advertisers. SSPs and ad networks have information regarding how impressions are served among advertisers/campaigns and associated ad responses across publishers. Ad exchanges have information about all winning and
losing bids from advertisers and inventory characteristics from publishers.

Limited data available to a single advertiser or publisher raises questions of how much/when/under which circumstances information should be shared between intermediaries and advertisers (and also publishers). For example, Sun et al. (2016) consider the timing of ad exchanges’ revelation to advertisers and find that hiding impression-level information at bid-time, but revealing at win-time, can increase the seller’s revenue, especially when the gap between the highest and second-highest bid is large. In addition, ad exchanges can consider providing bid landscape reports to advertisers (and/or publishers) to help them better predict the distribution of highest bids. Disclosing bid information to advertisers (or publishers) can range from revealing all bids and identity of the bidders to revealing simple summary statistics as quantiles. Advertisers may or may not consent to such disclosure and receiving more information about their competitors.

Intermediaries may have different data-sharing incentives from advertisers and publishers (Marotta et al. 2017). Rafieian and Yogaranarismihan (2018b) consider the data-sharing arrangements between the ad network and advertisers, and find that ad networks have the incentive to withhold information to increase advertiser competition and increase ad networks’ revenue. Resolving the information asymmetry between advertisers and publishers by the intermediary will further depend on the competition among intermediaries. For example, header bidding will change the incentives of intermediaries in data-sharing arrangements as competition among ad exchanges intensifies.

Another avenue for research is the market for data. Advertisers have different types of proprietary data in-house and often consider buying third-party data for additional user information. For example, Amazon has access to users’ browsing and purchase histories on its site, whereas P&G might have a harder time linking online users to their offline purchases. These different levels of user information raises the question of when it is beneficial for the advertisers to buy additional data and how the intermediary should price the information (Bhawalkar et al. 2014, Dalessandro et al. 2014). Related questions of interest pertain to the types of data third parties should collect that can best augment advertisers’ own proprietary data toward improving consumers’ ad experiences and privacy concerns. The information intermediaries share with advertisers or publishers can affect bargaining power among these players, so it would be desirable to address this problem as well.

### 6. Display Advertising Contexts

In addition to the more general display advertising considerations we note previously, various contextual factors are becoming more germane as technology in this space evolves. Specifically, we address mobile and social media in this section, and discuss video ads, transparency related issues, a brief history of display ads, and sponsored search advertising in the online appendix.

#### 6.1. Mobile

Various aspects of the display markets in mobile advertising warrant additional consideration owing to its rapid growth and unique ecosystem. First, search costs are higher (Ghose et al. 2012) and the information-carrying capacity is more constrained on mobile as screen sizes are smaller. As such, Bart et al. (2014) show that mobile display advertising is more effective for higher involvement and utilitarian (versus lower involvement or hedonic) products.

A second key point of difference is that mobile offers a wide new array of targeting variables that can be used for the purchase of impressions. Ghose (2017) enumerates context (e.g., Ghose et al. 2019b), location (e.g., Fang et al. 2015, Fong et al. 2015, Dubé et al. 2017), time (e.g., Danaher et al. 2015), saliency (e.g., Ghose et al. 2012), crowdedness (e.g., Andrews et al. 2015), weather (e.g., Li et al. 2017), historical shopping patterns/trajectory (e.g., Ghose et al. 2019a), social dynamics (e.g., Provost et al. 2015, Ghose et al. 2017), and tech mix as additional factors that moderate the efficacy of advertising. The value of this information for targeting is exemplified by Andrews et al. (2015), who demonstrate that users are responsive to marketing offers in crowded environments, and Luo et al. (2013) who detail the interaction between proximity to the focal venue and the time a promotion is received. An active area of research is how this rich information can be exploited in the mobile display ad ecosystem and the resulting implications for the players in this market.

A third unique aspect pertains to balancing advertising across mobile web and in-app. Although mobile apps consume nearly 90% of the time adults spend on their smartphone (including nonad supportive apps such email apps), there exists a greater diversity of publishers and audiences using mobile web. Accordingly, coordinating campaigns across apps and browsers requires additional consideration, though there is little research to date on understanding consumers’ interactions (e.g., in-app browsing) and advertising effectiveness across these channels.

Fourth, for the in-app advertising, there exist multiple platforms (e.g., iOS, Android). As users and apps differ
significantly across platforms, architecting advertiser campaigns across multiple platforms is an open issue. On the supply and intermediary sides, open questions include the role of app providers (who are often both publishers and advertisers, Lee et al. 2020), tying information across platforms or ad networks (for example, to enable in-app parallel bidding), and optimal ad-sequencing (e.g., in-app fading ads in Sun et al. 2017, Rafieian and Yoganarasimhan 2018a).

Finally, the concept of mobile advertising engagement is expanding and evolving. For example, information regarding consumers’ emotions from facial expressions (e.g., face unlock) or biometrics (e.g., watch) will likely affect advertising strategies. Likewise, concerns remain regarding how advertisers should think about audio advertising as consumers are increasingly turning to smart speakers for convenient access to news and entertainment.

6.2. Social Media
As with mobile advertising, social media advertising enables advertisers to tap into a richer set of variables for targeting users, such as the network structure itself and its users’ potential sharing and liking actions. Moreover, because almost all users are registered, social sites have a large set of user descriptors (Gordon et al. 2019). Finally, targeting in this context relies not only on the response to a single user, but the ripple effect of social influence (Zubcsek and Sarvary 2011, Bakshy et al. 2012, Tucker 2014a, 2016, Zhang et al. 2016a, Lee et al. 2018). Little research exists on bidding in ad markets where user valuations are interdependent.

Perhaps more germane is social advertising, which is dominated by Facebook. It is largely vertically integrated, so its ad buying network is vastly different from that characterized in Section 2. Not only does Facebook serve as a SSP, DMP, and ad exchange, but it is also a publisher. As a result, Facebook can more easily incorporate its own data in designing the auction mechanism and has greater control over balancing the needs of its viewern and the needs of its advertisers. For example, Facebook might accept lower bids in order to obtain higher reader engagement since showing ads that are not well aligned with reader interests can induce its viewers to leave the site. Trading off the relative weight of advertiser and user interests in determining the auction winner would be a beneficial avenue for research in both traditional (Mookerjee et al. 2016, Stouman and Bax 2017) and social media display markets.

7. Conclusion
The display ad market is economically substantial and rapidly growing. The growth is largely driven by RTB, in which many intermediaries exist to serve the technological needs of advertisers and publishers. The guaranteed selling channel and RTB have distinguishing characteristics, especially in terms of players involved, information available, and pricing mechanisms. Despite some industry experts’ belief that all ads will be bought and sold via RTB, the dual-selling channels will likely coexist in the long run as advertisers and publishers have incentives to leverage the distinctive features of both channels (Athey et al. 2016, Sayedi 2018).

In this article, the ecosystem of the display ad market, with dual-selling channels, is first outlined. Then, for each player involved (advertisers, publishers, and intermediaries), papers are reviewed across disciplines pertaining to each player’s decisions. Many open interdisciplinary research questions remain to be explored, especially in joint consideration of guaranteed and nonguaranteed selling channels, and we propose potential directions to stimulate fruitful research in this field.

Acknowledgments
The authors thank Garrett Johnson for helpful comments and suggestions.

Endnotes
3 The online appendix highlights similarities and dissimilarities of online display advertising from sponsored search advertising.
5 For a brief description of the history of online display advertising, please refer to the online appendix.
6 We use “nonguaranteed selling channel” and “RTB” interchangeably in this article.
8 Instead of guaranteeing a number of impressions to be delivered, contracts may alternatively be written to guarantee delivery dates or a share of impressions (Danaher et al. 2010) at a fixed price. Although less common, publishers may guarantee the number of clicks to be delivered based on cost-per-click (CPC).
9 The terminology “programmatic buying” should not be confused with nonguaranteed RTB. Programmatic buying consists of a wide range of technologies that automate the buying, selling, and matching of ad.
10 Depending on the publisher, negotiation is altogether precluded in programmatic direct to facilitate the automated contracting process. For example, the quoted price in Facebook’s “Reach and Frequency Buying” is a take-it-or-leave-it fixed price.
In their working paper version (Athey et al., 2014), section 7 explores publishers offering guaranteed versus nonguaranteed deals.

For example, some advertisers may value and want to target first-time home buyers. First-time home buyers in their 30s visiting *The New York Times* will be a smaller set than users in their 30s, and the visits of the former group on a given day will likely exhibit higher volatility.

According to the IAB, 65% of brand advertisers surveyed have fully or partially moved programmatic buying functions in-house. See https://www.iab.com/wp-content/uploads/2018/05/IAB_Programmatic-In-Housing-Whitepaper_v5.pdf.

Although in a different industry, Salz (2017) explains how the existence of intermediaries improves overall welfare by reducing incurred search costs and reallocating contracts to lower cost carriers in the New York City trade waste market.

In analyzing the second-price auction data from Microsoft Advertising Exchange, Celis et al. (2014) find that there is often a large gap between the highest and second-highest bid, and that the gap is bigger than the winning payment price on average.

According to eMarketer, the use of ‘first (second)’ price auction format increased (decreased) from 5.8% (75.1%) in December 2017 to 43.4% (32.8%) in March 2018. See https://www.emarketer.com/content/first-price-auctions-are-driving-up-ad-prices.

According to IAB, 65% of brand advertisers surveyed have fully or partially moved programmatic buying functions in-house. See https://www.iab.com/guidelines/.

Although in a different industry, Salz (2017) explains how the existence of intermediaries improves overall welfare by reducing incurred search costs and reallocating contracts to lower cost carriers in the New York City trade waste market.

In analyzing the second-price auction data from Microsoft Advertising Exchange, Celis et al. (2014) find that there is often a large gap between the highest and second-highest bid, and that the gap is bigger than the winning payment price on average.

According to eMarketer, the use of ‘first (second)’ price auction format increased (decreased) from 5.8% (75.1%) in December 2017 to 43.4% (32.8%) in March 2018. See https://www.emarketer.com/content/first-price-auctions-are-driving-up-ad-prices.

According to IAB, 65% of brand advertisers surveyed have fully or partially moved programmatic buying functions in-house. See https://www.iab.com/guidelines/.

Although in a different industry, Salz (2017) explains how the existence of intermediaries improves overall welfare by reducing incurred search costs and reallocating contracts to lower cost carriers in the New York City trade waste market.

In analyzing the second-price auction data from Microsoft Advertising Exchange, Celis et al. (2014) find that there is often a large gap between the highest and second-highest bid, and that the gap is bigger than the winning payment price on average.

According to eMarketer, the use of ‘first (second)’ price auction format increased (decreased) from 5.8% (75.1%) in December 2017 to 43.4% (32.8%) in March 2018. See https://www.emarketer.com/content/first-price-auctions-are-driving-up-ad-prices.

According to IAB, 65% of brand advertisers surveyed have fully or partially moved programmatic buying functions in-house. See https://www.iab.com/guidelines/.

Although in a different industry, Salz (2017) explains how the existence of intermediaries improves overall welfare by reducing incurred search costs and reallocating contracts to lower cost carriers in the New York City trade waste market.

In analyzing the second-price auction data from Microsoft Advertising Exchange, Celis et al. (2014) find that there is often a large gap between the highest and second-highest bid, and that the gap is bigger than the winning payment price on average.

According to eMarketer, the use of ‘first (second)’ price auction format increased (decreased) from 5.8% (75.1%) in December 2017 to 43.4% (32.8%) in March 2018. See https://www.emarketer.com/content/first-price-auctions-are-driving-up-ad-prices.
46 Of note, vertical integration considerations are extending to other contexts. AT&T's recent purchase of Time Warner Inc. is an attempt to bundle ad network (distribution) and content.

References


Johnson GA, Lewis RA, Reiley D (2016a) Location, location, location: Repetition and proximity increase advertising effectiveness. Working paper, Boston University, Boston.


