Auction Information Presentation and Monetization of Intellectual Property

Liying Yang, Joseph Derby, Mayukh Dass, and Yi Qian

April 17, 2026

Abstract

Markets for intellectual property are characterized by substantial information asymmetry and valuation uncertainty. This paper studies how the presentation and organization of patent information influence outcomes in patent auctions. We assemble a novel dataset of 1,743 intellectual property assets sold across eight patent auction events between 2006 and 2008 and exploit a natural experiment in which the organization and salience of information in auction catalogs changed during the sample period. Linking auction outcomes with the NBER Patent Database and COMPUSTAT, we examine how improved information presentation affects both the probability of sale and realized transaction prices while accounting for heterogeneity in patent characteristics and potential endogeneity in seller decisions. We find that improved information presentation significantly increases both sale likelihood and auction prices. The effects are stronger for patents with higher estimated prices, fewer backward citations, larger bundles, and longer remaining lifetimes. In addition, lower estimated prices increase the probability of sale but may reduce final prices, highlighting a strategic trade-off in seller disclosure. Finally, using supervised topic modeling of catalog descriptions, we identify technological terms that systematically predict auction success. The results provide empirical evidence on how information environments affect valuation in markets for intangible assets.

Keywords: Information Presentation, Intellectual Properties, IP Auctions, Natural Field Experiment, Valuation, Text Analysis, Topic Models

JEL codes: O34, D44, D83

Liying Yang is an assistant professor at the Shenzhen Audencia Financial Technology Institute, Shenzhen University, email: lyang@szu.edu.cn. ORCID: 0009-0000-6070-0285. Joseph Derby is a teaching assistant professor at the John Chambers College of Business and Economics, West Virginia University, e-mail: joseph.derby@mail.wvu.edu. Mayukh Dass is the James L. Johnson Chair in Business Administration and Professor of Marketing at the Rawls College of Business, Texas Tech University, MS 2101, Lubbock, TX-79409, tel: 806-834-1924; fax: 806-742-2199, e-mail: mayukh.dass@ttu.edu. Yi Qian is the Fred Siller Professor of Marketing and Behavioral Sciences at the Sauder School of Business, University of British Columbia, Henry Angus 567, 2053 Main Mall, Vancouver, BC V6T 1Z2, tel: 604-827-5154; fax: 604-822-4697, email: yi.qian@sauder.ubc.ca. ORCID: 0000-0002-9401-3384. Yi acknowledges the support by Social Sciences and Humanities Research Council of Canada [grants 435-2018-0519 and 435-2023-0306].

1. Introduction

Markets for intellectual property (IP) play an increasingly important role in modern innovation economies. IP assets—including patents, trademarks, and licensing rights—generate substantial financial value through transfers, licensing agreements, and litigation settlements. High-profile transactions such as Google's Patent Purchase Promotion (Harbert, 2015) and Huawei's acquisition of 90 smartphone-related patents from BlackBerry (Toulas, 2021) illustrate the scale and strategic importance of IP markets. Prior research shows that patents can influence firm valuation and capital market outcomes (Sood and Tellis, 2009; Simeth and Cincera, 2015). Despite their importance, however, markets for IP continue to exhibit substantial information asymmetry and valuation uncertainty, making the exchange of intangible assets particularly challenging. From a marketing perspective, IP assets also pose unique challenges: their intangible nature, domain-specific language, and lack of comparability make them difficult to evaluate and promote. In this context, how firms and intermediaries present and communicate IP information becomes a strategic marketing lever that may shape buyer interest and perceived value.

Information asymmetry has long been recognized as a central obstacle in IP transactions. Notably, strategic disclosures could lead to high transaction costs in the IP market and the presence of patent trolls. As summarized by Anton and Yao (2002), IP sellers face a fundamental dilemma that "Ideas are difficult to sell when buyers cannot assess an idea's value before it is revealed and sellers cannot protect a revealed idea" (p. 513). These frictions create substantial transaction costs and strategic behavior in IP markets. Various institutional mechanisms, such as patent pools, have been proposed, and standard-setting organizations or specialized intermediaries have emerged to address these challenges. As a new marketplace mechanism at the time, IP auctions aimed to reduce search costs for both buyers and sellers, as well as marketing costs for sellers (Hagiu and Yoffie, 2013). This environment encouraged sellers to waive confidentiality rights to increase

buyer engagement (Anton and Yao, 2008), creating an opportunity to study the impact of information presentation on IP transactions. We frame this as a marketing problem in terms of how the structure and salience of presented information delivered via auction catalogs and related materials can influence buyer behavior and commercialization outcomes for intangible assets.

The study is further enabled by a unique natural experiment in changing the organization and presentation of the listed patent information around the mid-point of the study period. We study how these changes influence outcomes in patent auctions, and whether improved information presentation facilitates IP monetization; and further investigate whether patent characteristics moderate the effects of information presentation on auction outcomes, and whether textual descriptions of patents provide predictive signals of auction success. We examine these questions through the lens of auction theory. In particular, the Linkage Principle predicts that auction revenue increases when sellers reveal more information about the value of the auctioned item (Milgrom and Weber, 1982; Milgrom, 2004). While the linkage principle has been extensively studied in theory, empirical tests in real-world markets remain relatively scarce, and most studies focus on the quantity of information disclosed rather than the structure and presentation of information.

IP valuation has never been easy. Even within auction markets, however, significant uncertainty remains in the valuation of intellectual property. Prior research on patent auctions has largely focused on patent attributes, such as citations, technological domains, and legal characteristics, as determinants of value (Fischer and Leidinger, 2014; Odasso et al., 2015; Sneed and Johnson, 2009; Vimalnath et al., 2018). While this literature advances our understanding of the technological drivers of patent value, relatively little is known about how the information environment in which patents are marketed affects buyer decision-making.

More broadly, a growing literature in empirical marketing and industrial organization demonstrates that the availability and structure of information can substantially influence economic behavior in markets characterized by uncertainty and search frictions. Research on online word-of-mouth and information diffusion shows that market signals shape consumer beliefs and demand (e.g., Godes and Mayzlin, 2004). Other works demonstrate how information availability and search frictions affect consumer decision-making and market outcomes (e.g., Bronnenberg et al., 2012; Seiler, 2013; Bollinger et al., 2011). More recent studies also highlight the role of information disclosure and market transparency in shaping market efficiency and price formation (e.g., Jin and Kato, 2006; Fradkin and Holtz, 2023). These findings suggest that the way information is organized and presented may influence how agents interpret signals and form valuations, particularly in markets involving complex and heterogeneous assets. Yet empirical evidence on how information presentation affects outcomes in markets for intellectual property remains limited.

The importance of information presentation is also well documented in the marketing literature. Past marketing literature suggests that the structure of information presentation plays a significant role in how consumers acquire and process information (Bettman and Kakkar, 1977; Biehal and Chakravarti, 1982). Theory of cognitive fit (Vessey, 1991) suggests that the efficiency and effectiveness of the decision-making process depend on the fit between the problem representation and the problem-solving task. Extending this theory to our context, we propose that when information presentation is consistent with the desired representation for patents, it will lead to transparency and reduction of perceived risk of purchase, thus leading to higher valuation (Kelton et al., 2010). Given that our context of analysis is patent auctions, we empirically examine the role of information disclosure and presentation from the auction theory perspective. Auction

theory stresses the importance of information. In particular, the linkage principle states that an auction house can increase its revenue by providing as much information as possible about the value of the auctioned items (Milgrom and Weber, 1982). Our study draws upon this theory and strives to understand the effects of information presentation, sellers' strategic information disclosure—such as setting estimated prices—and narratives underlying IP text descriptions on auction outcomes. We extend this theory by conceptualizing information presentation not only as a disclosure mechanism but as a marketing instrument that shapes perceptions, reduces cognitive friction, and guides buyer responses.

To study the research questions empirically, we gather a unique panel of field data on live patent auctions conducted by a specialized intellectual property auction house between 2006 and 2008, with a natural experiment of improved information presentation on the auctioned IPs. The dataset contains transaction outcomes for 1,743 intellectual property assets sold across eight auction events. Importantly, the setting provides a natural experiment in which the presentation of patent information in auction catalogs changed during the sample period. In particular, modifications to the information presentation to potential buyers were made following consultation with auction management by one of the authors, in cooperation with the Co-Founder and CEO of the auction house. The organization and salience of catalog and IP information presentation were improved for three of the eight auction events. This change allows us to compare auction outcomes before and after the modification while holding other institutional features of the marketplace constant.

During the process of IP purchase, sellers provide buyers with comparative information for IP valuation that is not publicly accessible and help them to make valuation judgments. Such information is typically delivered in the form of information sheets and catalogs that provide

summary information, access to the seller's confidential due diligence materials ("data room") for detailed information, and private meetings with the sellers to clarify any questions developed during the due diligence process. These information sheets include a number of salient cues about the offering and its value that are presented in some form. We define "better" information presentation as instances when they are "organized" in their presentation in comparison to situations where the information presentation of the qualitative drivers of IP value is "disorganized." In our analysis, we first qualitatively analyze these changes and then test the main and moderated effects of information presentation on IP value using these transaction data linked with the NBER Patent Database and COMPUSTAT Database. We examine the influence of IP features that are strategically determined by sellers on auction outcomes as well, addressing potential endogeneity through a copula correction approach.

Our empirical analysis yields several findings. First, improvements in the organization and salience of patent information significantly increase both the probability that a patent sells and the realized auction price. Second, these effects vary systematically with patent characteristics: the impact of improved information presentation is stronger for patents with higher estimated prices, fewer backward citations, larger patent bundles, and longer remaining lifetimes. Third, we document a trade-off associated with sellers' estimated prices: lower estimates increase the likelihood of sale but are associated with lower realized prices. Finally, textual analysis of catalog descriptions reveals that certain technological topics and terms are systematically associated with successful auction outcomes, suggesting that narrative signals contained in auction materials influence bidder perceptions and valuation.

IP monetization plays a significant role in the value of firms, but how marketing plays a role in such a process has not yet been studied. In that respect, we make substantive and

methodological contributions to the literature. First and foremost, our participation and engagement with the auction house resulted in a unique natural field experiment, especially resulting from direct consulting on the information presentation policy. We compile a novel comprehensive dataset with patent panels before and after the policy change as detailed in the previous paragraphs.

Second, we provide one of the first empirical analyses of how information presentation affects outcomes in patent auction markets, extending the intellectual property valuation literature beyond technological attributes of patents to include the information environment in which they are marketed. Although our study is limited to the investigation of information presented to buyers, our paper brings the importance of marketing to the limelight. Extending the IP valuation literature (Harhoff et al., 2003), our study reveals that enhancing information clarity in auction materials boosts transaction outcomes and the strength of this effect varies systematically with core IP characteristics. Specifically, with organized information presented, patents with higher sellers' estimated prices, fewer backward citations, more patents pooled in a lot, and later expiration dates are more likely to sell and have a higher winning bid. We also provide empirical evidence for the trade-off faced by sellers when setting estimated IP market prices: while lower estimated prices attract more bidders and increase sale likelihood, they may also suggest lower final bids.

Third, we provide new empirical evidence related to the Linkage Principle (Milgrom and Weber, 1982) by showing that improvements in the clarity and organization of information presented to bidders increase both the probability of sale and realized auction prices. By exploiting a natural experiment in catalog presentation, our analysis isolates the causal impact of information presentation on auction outcomes. This is also one of the few real-world applications that provide a direct empirical test of the linkage principle.

Last but not the least, we contribute methodologically by combining auction transaction data with textual analysis of patent catalog descriptions using supervised topic modeling, allowing us to identify technological topics and terms that systematically predict auction success. Our adoption and improvisations of a supervised topic model enable us to identify influential topics and terms in auction catalogs that most strongly affect sale probability, either positively ("pro") or negatively ("passive"). For example, terms like "information server" and "interactive television" rank among the most pro-sale words (note that our data is over the period 2006–2008). These findings offer practical insights for IP managers to enhance search and organization strategies for future sales.

More broadly, our results contribute to the literature on information frictions and market design, highlighting how the organization and presentation of information can influence valuation in markets for complex and heterogeneous assets.

2. Literature and Frameworks

Our paper is broadly related to the literature on information and presentation associated with IPs and auctions that can be categorized into four areas – information disclosure related to IPs, IP economic valuation, IP transactions and market, and impacts of information presentation on auctions – summarized in Table 1.

2.1 IP Valuation Uncertainty and Information Presentation

Efficient valuation and exchange of IP are vital to innovation markets but are often impeded by informational and transactional frictions. A major source of these frictions is the inherent difficulty of valuing IP, driven by the specificity of patents and the lack of comparable market data. As a result, valuations are often costly, time-consuming, and uncertain, complicating

Table 1: Literature Related to Information Presentation Associated with IPs and Auctions

Topic	Studies
1. Information disclosure	
(1) Impacts on IP	Fan et al. (2016), Hong et al. (2021), Kang et al. (2024)
(2) IP disclosure strategy	Anton and Yao (2002, 2008), Jeon (2019)
(3) Impacts of IP disclosure	Hedge and Luo (2018), Lück et al. (2020), Furman et al. (2021), Beyhaghi et al. (2023), Reuer et al. (2025)
2. IP economic valuation	
(1) Observational data	Bessen (2008), Chen et al. (2019), Choi et al. (2020),
(2) Survey data	Harhoff et al. (2003), Gambardella et al. (2008)
(3) Merged with firm data	Sood and Tellis (2009), Wu et al. (2015), Kogan et al. (2017), Testoni (2022), Arts et al. (2023)
(4) Real transaction data: auctions, licensing, and transfers	Sneed and Johnson (2009), Gambardella et al. (2007), Serrano (2010), Fischer and Leidinger (2014), Odasso et al. (2015), Caviggioli and Ughetto (2016), De Marco et al. (2017), Vimalnath et al. (2018), Arora et al. (2022)
3. IP transaction and market	
(1) Other analyses related to IP transactions	Galasso et al. (2013), Caviggioli et al. (2017), Figueroa and Serrano (2019), Kwon et al. (2022)
(2) IP market	Arora and Gambardella (2010), Dushnitsky and Klueter (2017), Serrano (2018)
4. Evaluating impacts of information presentation on auctions	
(1) Presentation	Lacetera et al. (2016)
(2) Information quality	Kagel and Levin (1986), Tadelis and Zettelmeyer (2015), Aubry et al. (2023)

Notes: The extensive literature on innovation production and IP scientific valuation is not included, as they are outside the scope of this paper.

the monetization of IP rights. These challenges have resulted in the emergence of a secondary market for IPs (Chesbrough, 2006; Arora and Gambardella, 2010; Caviggioli et al., 2017). Particularly, different degrees of search costs, ambiguity about the underlying knowledge and its applications, and expropriation concerns affect which industry can be best served by IP platforms (Dushnitsky and Klueter, 2017; Serrano, 2018). Moreover, information disclosure decisions related to IP are often strategically set by firms facing different tradeoffs (Anton and Yao, 2008; Hong et al., 2021).

From a marketing perspective, these informational frictions mirror long-standing challenges in how complex, intangible, and unfamiliar products are presented to potential buyers. In such high-uncertainty contexts, the design and structure of information become central to

shaping perceived value, reducing decision risk, and facilitating exchange. Against this backdrop, by leveraging a unique natural experiment setup, in which the timing of improvements in information presentation is exogenous, our paper provides an empirical examination of how IP auctions can better facilitate gains from trade.

To address the challenges, extensive research has explored the role of information provision in shaping the value of IP, both scientific and commercial. Building on this literature, our paper examines whether and how information provision influences IP values. Policies mandating IP disclosure, such as the American Inventors Protection Act, mitigate information costs for different agents, for example, improving the accuracy of analyst forecasts (Beyhaghi et al., 2023), reducing duplicated inventions (Lück et al., 2020), as well as facilitating licensing agreements (Hedge and Luo, 2018; Jeon, 2019) and technology acquisition contracts (Reuer et al., 2025). The exchange of information, when enabled, also benefits knowledge creation (Kang et al., 2024). Our paper complements this strand of the literature: we show that improved information presentation, through the joint efforts of the auction house and sellers, facilitates IP transactions and thus makes greater gains from trade realized. Importantly, we extend this stream by demonstrating that not only the content of the disclosure, but the form and structure of the information itself (i.e. information design) play a pivotal role in commercialization success.

Our paper also relates to the literature on how information presentation affects consumer decisions and purchasing behavior. In marketing, information is defined as "data that have been given a structure—placed in context, and thus endowed with meaning" (Glazer, 1991). Research shows that the structure and format of information strongly influence how consumers acquire, process, and evaluate it, often favoring formats that minimize cognitive effort (Bettman and Kakkar, 1977; Biehal and Chakravarti, 1982; Jarvenpaa, 1989). Such cognitive simplification

effects are especially relevant in B2B and technology settings, where decision-makers face high uncertainty and complexity. Given humans' limited information processing capacity, people tend to choose the path of least mental effort when making decisions (Einhorn and Hogarth, 1981). Studies on information overload further suggest that consumers can typically compare five to ten attributes before experiencing overload (Malhotra, 1982), and that high-quality information enhances both information use and confidence in purchasing decisions. Overall, the literature indicates that decision makers prefer information that is succinct, structured, and meaningful (Todd and Benbasat, 1992). Our paper contributes to this ongoing discussion by providing empirical evidence that enhanced information presentation shapes how IP buyers perceive key drivers of IP valuation, thereby helping to mitigate information costs for both buyers and sellers. This insight highlights the marketing function of information design as an underutilized lever in markets for intangible assets.

Despite efforts to improve information environments, uncertainty in IP valuation remains a major challenge. Motivated by this, we explore whether machine learning techniques can help address this issue. Prior studies have demonstrated the value of developing new metrics for assessing the economic value of IP. For example, Kogan et al. (2017) use stock market responses to patent news as a novel measure, while Sood and Tellis (2009) highlight long-term stock gains from innovation announcements. More recently, Arts et al. (2021) show the advantages of applying machine learning methods to assess the scientific value of patents, and Testoni (2022) applies similar techniques to examine spillover effects in mergers and acquisitions. Building on these advances, we employ ML-based text analysis to extract features relevant to IP sales in auctions and to disentangle their effects from those of patent characteristics discussed below.

2.2 Drivers of IP Valuation

A variety of IP valuation drivers have been identified in the literature. The value drivers can be grouped into two broad categories, namely, product characteristics and seller and buyer characteristics. The product characteristics include a patent's number of forward citations, if it has been licensed, its age, and the number of claims of the patent; and the seller characteristics include whether the seller is a company or an individual. As information on buyers is not available in our data set, we discuss product and seller characteristics in detail as follows. See Caviggioli and Ughetto (2016) for the impacts of buyer characteristics.

Product characteristics. The number of forward citations has been shown as one of the most crucial value drivers in the setup of IP auctions (Sneed and Johnson, 2008; Fischer and Leidinger, 2014; Odasso et al., 2015) or other patent datasets (Harhoff et al., 2003; Gambardella et al., 2008; Serrano, 2010). Patent age is also a relevant feature with a possibly nonlinear relationship with patent commercialization potentials (Vimalnath et al., 2018; Choi et al., 2020). Other important patent characteristics include the patent's family size, patent scope (Gambardella et al., 2007), patent transaction history, etc.

Seller characteristics. Existing literature suggests that patents from small companies or private inventors are less valuable than those from larger firms (Bessen, 2008; Sneed and Johnson, 2008) but are more likely to be traded (Gambardella et al., 2007; Arora et al., 2022). The firm size's effect, combined with technological and geographical proximity of existing knowledge stocks, is also explored by Figueroa and Serrano (2019) and Galasso et al. (2013). The heterogeneous effects of technological and geographical proximities are also studied by De Marco et al. (2017), Figueroa and Serrano (2019), and Kwon et al. (2022).

Industry. IP values realized in the market vary across industries (Serrano, 2010), reflecting

different norms of licensing, transaction costs, and perceived enforceability. Platform and auction dynamics may also interact with these industry-level effects, as some domains are more conducive to third-party validation or cross-licensing opportunities than others (Dushnitsky and Klueter, 2017). Therefore, we control for these factors in our model using industry-fixed effects.

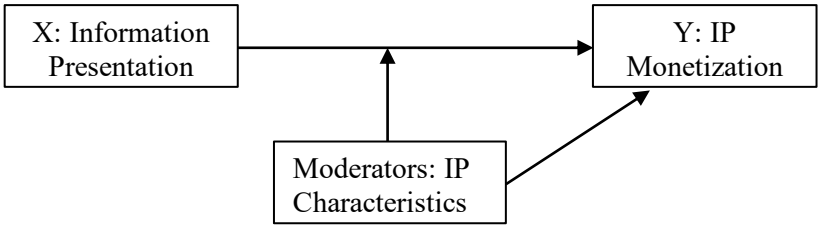
Marketing relevance. Recognizing that information overload may cause buyers to undervalue detailed but poorly organized information provided by sellers, our paper provides empirical evidence for both the direct effects of key value drivers and how these effects vary depending on whether information is presented in an organized or disorganized format. From a marketing standpoint, this moderating role of information presentation suggests that the presentation of value signals, such as citations or estimated value, can either amplify or obscure their influence, depending on how they are framed and structured. In addition, we examine the roles of sellers' strategic decisions, including signaling through estimated IP values and listing potential licensees, while accounting for potential endogeneity of these factors using the two-stage copula control function estimation (2sCOPE) approach by Yang, Qian, and Xie (2024). Notably, the main effect of information presentation or its interaction with key patent value drivers have not been explored in prior studies, which did not recognize the unique natural experiment leveraged in our analysis.

2.3 Empirical and Conceptual Frameworks of This Paper

In this paper, the outcome variables of interest include both the probability of an IP sale and the sale price as the winning bid, based on the research framework illustrated in Figure 1.

Since buyers lack access to comparative information for IP valuation, a central role of the auction house is to provide essential information to support valuation and facilitate the sale. The

Figure 1: Research Framework



Notes: IP characteristics include IP structured characteristics, seller characteristics, and IP latent text structure.

auction catalog, as the primary information source for potential buyers, serves as a critical marketing tool. Well-structured and salient catalog content can reduce perceived risk, improve comprehension of value drivers, and enhance overall buyer confidence. We conjecture that the catalog's presentation provides valuable cues about the IP offering, with well-organized, salient information aligned with IP value drivers leading to higher auction outcomes. Conversely, disorganized information that lacks alignment with key value indicators may negatively impact auction results. Moreover, the effect of information presentation on IP value may differ across various types of IP. In addition to analyzing how structural characteristics such as citations shape this relationship, we extract latent textual features using a topic model to capture the underlying content structure. This allows us to investigate how deeper semantic information influences auction outcomes and interacts with presentation quality to influence buyer responses.

We use linkage theory and cognitive fit theory as the foundation for our conceptual framework. Cognitive fit theory links lack of proper information presentation format to ineffective decision making and high perceived risk (Vessey, 1991). Linkage theory explains the relationship between the types of information disclosed by the seller and the number of bidders attracted to the auction (Milgrom and Weber, 1982). Information disclosure in linkage theory considers various policies of reporting private information by sellers. According to the linkage theory, sellers could

increase the average auction receipts by releasing private information about the item being sold that correlates with bidders' valuation and by linking prices to exogenous variables that are affiliated with bidders' private information. Extending these theories into a marketing context, we posit that the format, salience, and clarity of information which is core to marketing communications, function not only as informational cues but also as persuasive devices that shape buyer engagement and perceived value. In the Appendix, we extend the framework and analytically describe how information presentation may increase auction revenue for the sellers.

Prior research based on this theory has examined how sellers can strategically release information and select auction types to increase revenue, including Kagel and Levin's (1986) experimental auctions and Tadelis and Zettlemeyer's (2015) field experiment on randomly disclosed quality information in wholesale car auctions. Lacetera et al. (2016) further show that auctioneer performance impacts car auction outcomes. Extending this literature, we study information presentation in IP auctions. Aubry et al. (2023) find that ML-based valuations could correct persistent prediction errors in art auctions; similarly, we use a supervised ML model to uncover latent text features driving patent sales, such as the underlying segmented technologies as in Chen et al. (2019). These contributions advance the literature on marketing of intangible goods by offering one of the first empirical tests of how structured information design, aligned with both cognitive fit and signaling theory, can influence buyer behavior and auction performance in B2B markets.

3. Background and Data

3.1 Natural Experiment

IP auctioned by Ocean Tomo undergoes an intensive screening process using proprietary analytics aimed at identifying high-quality IP assets that are likely to be sold at auction. For

example, in the first patent auction held in 2006, Ocean Tomo selected 430 patents for auction out of over 1,200 patents submitted by sellers (Malackowski, 2006). Similar to auctions in many other product categories, Ocean Tomo made their catalogues approximately 2 months prior to the auction, although it can be much shorter (even as little as one week) for some individual lots that were added later.

Ocean Tomo organized 10 live IP auctions between April 2006 and June 2009, which generated a lot of industry and media buzz. There has even been academic optimism (e.g., Gans and Stern, 2010) regarding the potential for bringing liquidity to the IP market via platforms. Due to the financial crisis, the last two auctions saw rather stagnant sales. The total value of transactions through Ocean Tomo's 10 IP auctions was $114.6 million (Jarosz, 2010). The average sales-to-listings ratio over all 10 auctions was reportedly 38%, while the Spring 2009 auction only sold 6 out of 85 lots listed. We therefore only keep the 8 auctions prior to the outburst of the financial crisis in our analyses.

In this study, we are privileged to exploit a *natural field experiment* to study how information presentation affects IP valuations through one of the authors' continuous contacts with the Ocean Tomo auction house. After observing one of the 2007 IP auctions in Chicago, the CEO of Ocean Tomo asked the author for advice during a conversation. A set of recommendations was made to the auction house for improving information presentation in the IP advertisements, auction catalogs, and live auction projections. In particular, suggestions were made and adopted to make the IP flyers and catalogs less cluttered, to better present key information of each IP, to add dual-screen projection in live auctions with one screen displaying key information of the IP being auctioned and the other screen displaying the sequence of live bids, etc. The auction house adopted the recommendations for the subsequent auctions in 2008 and onwards. The IP auction outcomes

and characteristics data from the 8 auctions enable valuable before-after tests of this unique field experiment. The changes in the presentation styles and information disclosure of the subsequent auctions after the natural experiment intervention are evident in the data. The lead time of sending out catalogs did not see significant changes.

In each live auction, an auctioneer takes bids for each lot, which could be a single IP asset (patent, copyright, trademark, or domain name right) or a bundle of such assets. The lots were sold to the highest bidder as long as the highest bid exceeded the seller's reserve price if present. Ocean Tomo collected fees from both sellers and buyers: sellers paid a listing fee, and bidders paid registration. In some auctions, Ocean Tomo also received a 10% buyer's premium and a 15% seller's premium when transactions were successfully closed (Hagiu and Yoffie, 2013).

3.2 Qualitative Exploration of Auction Catalogs

We first qualitatively examine eight auction catalogs that were consecutively held in the US and in Europe over 2006-2008. This data represents the publicly available data that was released by the auction house and the marketing information that is known to be common to all potential bidders.

In general, we found that information presentation within each auction catalog was consistent. The IP was presented as lots, consisting of either single IP entities or combinations of IP entities. The vast majority of the IPs were in the form of US patents; relatively few were in other forms such as trademarks, domain names, copyrights, and foreign patents. On the other hand, we found that information presentation across auction catalogs was not consistent in terms of organization and type. For example, IP information in early 2006 auction catalogs were not grouped systematically by technology area nor auctioned in sequence. IP information in 2008 auction catalogs was grouped and color-coded by technology area and auctioned in sequence.

Information categories and presentation salience also changed over time across auction catalogs. The most salient information in early auctions was the lot number, area of technology, patent number, assignee/lead inventor, and title. In later auctions, the most salient information was the area of technology and the title of the lot, where the seller of the lot, the featured IP asset, related IP, and expected value were clearly listed as well.

We tracked how auction catalog information presentation changed over time. For each auction catalog, we listed the information provided for each lot and categorized how information was presented for each intellectual property. We summarized the information presentation for each auction by major category (Lot number, Patent Number, Title, Author, Seller, Assignee, Number of Claims, Sampled Forward Citations, etc.) Our measure for organization is a combination of salience and importance of information. Salience is a combination of emphasis, color, and position on a page. The most salient information has the highest emphasis, is color coded, and is placed at the top of the page. The most important and relevant information is the information about what is being purchased – the drivers of the IP valuation such as citations and clear information about what is being purchased – the title of the lot, and from whom it is being purchased – the seller of the lot. We defined organized information as presenting important information saliently and disorganized information as presenting unimportant information saliently or important information unsaliently.

Based on auction theories and information-disclosure concepts, we develop our categorization schemes for important and unimportant, salient and unsalient, and organized and disorganized information presentation. We define auction catalogs that displayed important information saliently as "organized" and auction catalogs that displayed unimportant information saliently or important information unsaliently as "disorganized." We hired three independent

judges to classify the auction catalogs. We had the judges rate the descriptions, presentation, organization, and clarity of each catalog on a 7-point Likert scale, and also provided an open-ended question for them to comment on anything they noticed about the information presentation. The judges' ratings and codings indicate that auction catalogs before Spring 2008 were "disorganized" and auction catalogs from Spring 2008 forward were "organized". The inter-rater agreement was found to be 88.7% (Cohen, 1960). In the following, we formally conduct a quantitative analysis of the effects of improved information presentation.

3.3 Auction Information: Sample and Structured Variables

We develop an extensive data set constructed from the auction catalogs and auction results and include all the public information provided by the auction house. The information includes the location of the auction, the date of the auction, a description of each lot, the individual patents in each lot, the title of each patent, the inventors, the assignees, the sellers, the technology category, potential licensees of the technology, the number of claims, the number of forward citations, the date of patent issuance, the date of the auction, the range of value and the expected value of the lots, the auction outcome of the lot, and the price of lots sold. We calculate the age of the IP at the time of auction by subtracting the date of patent issuance from the date of the auction. We count the number of patents offered in each lot.

The US Patent Classification System (USPCS) classifies utility patents based on common subject matter, with a class generally delineating one technology from another. Analyzing inventive activity by industry requires a concordance of the patents to industries that would practice the technologies. We map the patent classification of each patent in our data set to the North American Industry Classification System (NAICS) using the concordance method established by the USPTO. NAICS was adopted in 1997 to replace the Standard Industrial

Classification (SIC) and is the present standard used by Federal statistical agencies in classifying business establishments.

The sample used in our analysis comprises 1,377 patents granted in the U.S., for which we are able to retrieve the information on forward citations (weighted by truncation as of 2006) from the NBER patent database and backward citations from the Google patent database.

Table 2 presents summary statistics for the outcome variables and patent characteristics before and after treatment (i.e., the natural experiment). We find that 21% more of the offered patents are sold after the treatment, and in addition, among the patents that are sold, the average sale price is 30% higher compared to the pre-treatment average. Figure 2 presents two empirical power functions of Log(Price) for patents pre- and post-treatment. The left sub-plot includes both sold and unsold patents (with Log(Price) = 0 for unsold patents), while the right sub-plot focuses only on sold patents. As can be seen in the left sub-plot, a larger fraction of patents remained unsold in the pre-treatment period, as indicated by the higher density near zero. Additionally, post-treatment patents were sold at relatively higher prices, as seen from the greater density in the right tail of the distribution of sold patents only.

These findings motivate us to examine the causal effects of improved information presentation on IP commercialization, as well as how it shapes the impact of patent characteristics—referred to as value drivers—on commercialized patent values. We categorize the patent characteristics into three groups: (i) featured characteristics, which require sellers to make judgments; (ii) citation metrics, which are crucial value drivers for patent scientific values; and (iii) other characteristics.

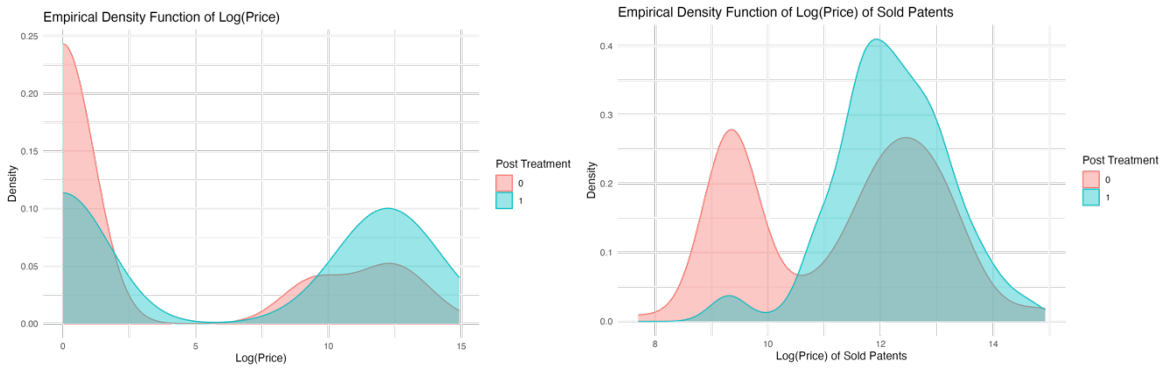
We notice that after treatment, on average, sellers set a lower estimated (targeted) sale price with fewer patents pooled in a lot. Similar to the reserve price, setting a lower estimate may attract

Table 2: Summary Statistics of Outcome Variables and Patent Characteristics

	Outcome variables			Featured characteristics (Features)			Citation metrics (Citations)			Other characteristics (Others)		
	Sold	Price	Log(Price)	Estimate	Log(Estimate)	No. potential licensees	Forward citations	Weighted forward citations	Backward citations	No. patents in a lot	Branded	Patent age
Before treatment (n=1,013)												
Mean	0.29	72983	3.27	1550071	13.43	2.87	10.61	20.90	18.42	12.59	0.74	6.88
SD	0.45	268827	5.21	1755907	1.49	0.87	12.49	35.02	20.67	14.19	0.44	4.39
Min	0.00	0	0.00	10000	9.21	0.00	0.00	0.00	1.00	1.00	0.00	0.18
Q25	0.00	0	0.00	300000	12.61	2.00	2.00	2.51	6.00	3.00	0.00	3.30
Median	0.00	0	0.00	750000	13.53	3.00	7.00	10.19	12.00	8.00	1.00	6.21
Q75	1.00	11000	9.31	2000000	14.51	3.00	14.00	24.47	22.00	17.00	1.00	9.91
Max	1.00	3025000	14.92	5000000	15.42	6.00	87.00	408.05	166.00	59.00	1.00	20.32
After treatment (n=324)												
Mean	0.50	163257	6.14	504062	12.62	2.68	16.36	27.44	13.55	2.78	0.47	7.17
SD	0.50	316553	6.15	601995	1.06	0.69	22.88	44.41	10.66	2.64	0.50	3.94
Min	0.00	0	0.00	10000	9.21	0.00	0.00	0.00	1.00	1.00	0.00	0.34
Q25	0.00	0	0.00	150000	11.92	2.00	3.00	0.00	6.00	1.00	0.00	3.77
Median	1.00	11000	9.31	310000	12.64	3.00	9.00	13.54	11.00	2.00	0.00	6.85
Q75	1.00	193875	12.17	550000	13.22	3.00	16.00	34.39	17.50	3.00	1.00	9.63
Max	1.00	2414104	14.70	4000000	15.20	6.00	142.00	307.29	71.00	12.00	1.00	18.41

Notes: “Sold” is a dummy variable indicating whether a patent was sold in an auction or not. “Price” represents the sale price, and “Log(Price)” is the natural logarithm of the sale price for sold patents and set to zero for unsold patents. “Estimate” refers to the seller’s estimated sale price, which can be equal to or higher than the reserve price (if applicable), and “Log(Estimate)” is the natural logarithm of the estimated value. “No. potential licensees” represents the number of potential licensees. Since the first two auctions did not list potential licensees, summary statistics for the pre-treatment period are based on 444 patents from Auctions 3, 4, and 5. “Forward citations” refers to the number of forward citations from prominent companies such as Apple, listed on the auction catalog. “Weighted forward citations” refers to forward citations from the NBER patent database, adjusted with truncation weights as of 2006. “Backward citations” refers to the number of the patent’s references from the Google patent database. “No. patents in a lot” refers to the number of patents bundled in a lot. “Branded” is a dummy variable indicating whether the seller is a branded company. “Patent age” measures the number of years between the patent’s issue date and the auction date. The data for these variables are from patent auction catalogs unless otherwise specified.

Figure 2: Empirical Density Functions of Log(Price) for All Patents (Left) and Sold Patents Only (Right) Pre- and Post-Treatments



more potential buyers, potentially benefiting the sellers. This might also result from an average decrease in the number of patents in a lot. As shown in our later analysis, it is crucial to cap the patent portfolio size to enhance its liquidity. In the first two auctions, potential licensees were not listed. Excluding these auctions, the average number of potential licensees before treatment was slightly higher than the post-treatment average. For patent citation metrics, we obtained the number of forward citations from the catalog ("Forward citations"), which represents a subset of the truncation-weighted forward citations from the NBER database ("Weighted forward citations"). After treatment, patents in these later auctions received citations and had, on average, 54% more sampled citations listed in the catalog, which may signal higher IP value. Also, they had, on average, 26% fewer backward citations, which could indicate greater originality.

A summary of the number of patents in different technology classes of the NAICS is provided in Table 3. The top four categories with the highest number of patents offered in auctions, both before and after treatment, are "Communications Equipment," "Computer and Peripheral Equipment," "Semiconductors and Other Electronic Components," and "Navigational, Measuring, Electromedical, and Control Instruments." The fifth category shifted from "Other Chemical Product and Preparation" before treatment to "Medical Equipment and Supplies" after treatment, reflecting a change in technology classes following treatment.

Table 3: Number of Patents in Different NAICS Industry Categories

NAICS ID	NAICS category	Before		After	
		Sold	Unsold	Sold	Unsold
1	Food	0	11	0	0
2	Beverage and Tobacco Products	0	2	0	0
3	Textiles, Apparel and Leather	1	9	0	0
5	Paper, Printing and support activities	0	3	0	0
7	Basic Chemicals	2	58	0	0
8	Resin, Synthetic Rubber, and Artificial and Synthetic Fibers and Filaments	3	4	0	0
9	Pharmaceutical and Medicines	0	6	0	0
10	Other Chemical Product and Preparation	10	79	0	2
11	Plastics and Rubber Products	2	21	0	1
12	Nonmetallic Mineral Products	0	6	0	0
13	Primary Metal	0	2	0	0
14	Fabricated Metal Products	1	35	0	2
15*	Machinery	15	55	5	8
17*	Computer and Peripheral Equipment	61	71	59	48
18*	Communications Equipment	40	138	40	49
19*	Semiconductors and Other Electronic Components	38	94	13	13
20*	Navigational, Measuring, Electromedical, and Control Instruments	53	60	19	23
21*	Other Computer and Electronic Products	6	17	7	3
22*	Electrical Equipment, Appliances, and Components	16	49	4	1
24	Motor Vehicles, Trailers and Parts	36	7	1	0
25	Aerospace Product and Parts	1	1	0	0
26	Other Miscellaneous	6	1	0	0
29	Other Transportation Equipment	0	11	0	0
30*	Medical Equipment and Supplies	11	11	15	11
Sub-sample*	1,053	240	495	162	156
Total	1,377	302	751	163	161

Notes: For robustness check, we also construct a sub-sample that only includes NAICS categories with sold patents in both the pre-treatment and post-treatment periods.

After describing how we abstract the textual characteristics from auction catalogs, we illustrate our empirical strategies in Section 4 to address endogeneity from selection bias in terms of which patents were chosen by sellers and/or the auction house over different auctions.

3.4 Textual Characteristics

As the improvement in presentation format is well documented, it is likely that sellers also

provided more attractive content to introduce the offered patents. To analyze the catalog texts for each lot, we first consider cosine similarity (ranging between -1 and 1) as a straightforward measure of semantic similarity based on the text-word frequency matrix, which helps us understand how closely related the catalog texts are in terms of meaning. Since this measure only involves the descriptive text-word frequency, we further employ the unsupervised Topic-SCORE approach by Ke and Wang (2024) to identify and estimate the latent topic structure that generates the texts. The Topic-SCORE approach imposes the anchor word assumption, that is, there exist signature words that uniquely define each topic, which facilitates the discovery of meaningful and distinctive topics. It can tackle severe word frequency heterogeneity, which is desirable for analyzing patent texts that inevitably comprise various technical jargons. It also outperforms latent Dirichlet allocation (LDA) models with smaller errors in simulations, regardless of whether the data is generated under the anchor word assumption or the LDA model. Given these desirable properties of the estimated topic-word distribution, we calculate topic weights for each text document to provide a breakdown of the topics covered. This allows us to capture variations in the content presented.

Table 4 presents summary statistics for word count, semantic similarity, and estimated document weights across three distinct topics.[1] The top 20 representative words in Table 5 indicate that these topics can be interpreted as "Computers and Communication," "Chemicals," and "Semiconductors and Instruments." These topics are broader in scope than the technology classes and could serve as overreaching themes for positioning different auctions, potentially attracting

[1] When three topics are specified, the estimated topic models are stable across different seeds used in the singular value decomposition process in the Topic-SCORE approach. However, when the number of topics is set to four, the Topic-SCORE approach fails to identify a meaningful fourth topic. Instead, the representative words for this topic consist only of the most frequently occurring words (not chosen by other topics) in the text, without reflecting a coherent theme. Hence, we stay with the choice of three topics.

buyers with relevant business experience and interests in these areas. Overall, the texts contain fewer words after treatment, probably due to the reduced number of patents in an auction lot. The semantic similarity of catalog texts is comparable before and after treatment. We find that the degree of semantic similarity is relatively low, around 0.07–0.08 (with a possible range between 0 and 1), which is expected due to the technical nature of patent descriptions. The estimated document-topic weights reveal a shift in the technology categories of patents offered at auction. Before treatment, chemical-related patents were more prevalent, whereas after treatment, there was an increase in semiconductor- and instrument-related patents, which were popular and trending in 2008. This pattern aligns with the summary results presented in Table 3.

Table 4: Summary Statistics of Variables from Textual Analysis

	Document word frequency (Word)				Document topic weights (Topics)		
	Word count	Log(Word count)	Semantic similarity	Scaled semantic similarity	Weight on topic 1	Weight on topic 2	Weight on topic 3
Before treatment (n=1,013)							
Mean	1296	6.83	0.07	3.31	0.15	0.68	0.18
SD	1147	0.83	0.02	1.01	0.20	0.27	0.18
Min	54	3.99	0.02	0.99	0.00	0.00	0.00
Q25	512	6.24	0.05	2.44	0.00	0.48	0.00
Median	924	6.83	0.07	3.24	0.09	0.67	0.16
Q75	1545	7.34	0.09	4.10	0.22	0.97	0.29
Max	4814	8.48	0.13	5.71	1.00	1.00	0.83
After treatment (n=324)							
Mean	741	6.50	0.08	3.59	0.20	0.47	0.33
SD	397	0.44	0.02	0.94	0.12	0.15	0.13
Min	259	5.56	0.03	1.25	0.00	0.00	0.00
Q25	486	6.19	0.07	2.95	0.14	0.39	0.25
Median	594	6.39	0.08	3.47	0.18	0.46	0.31
Q75	909	6.81	0.10	4.36	0.24	0.54	0.41
Max	2120	7.66	0.13	5.75	0.64	1.00	0.85

Notes: "Word count" represents the total number of words in a catalog text document. "Log (Word count)" denotes the logarithm of word counts. "Semantic similarity" is calculated as the average cosine similarity between the focal patent's document-word frequency vector and those of other patents in the sample. "Scaled semantic similarity" is calculated by standardizing the level of semantic similarity using its standard deviation across all patent texts. "Document topic weights" are estimated based on the Topic-SCORE approach by Ke and Wang (2024) with three topics.

Table 5: Top 20 Representative Words for Each Estimated Topic

Topic 1: Computers and Communications		Topic 2: Chemicals		Topic 3: Semiconductors and Instruments	
Word	Count	Word	Count	Word	Count
payload	89	alkyl	75	frame	120
routing	87	bleaching	64	stream	104
header	76	peracid	52	token	85
embedded	44	atoms	44	word	77
signaling	43	str1	43	buffer	73
versions	40	bleach	40	decoder	71
propagate	30	precursors	25	stages	70
replicated	22	quaternary	25	encoded	66
encompassing	19	saturated	24	words	66
survivability	19	nitrile	21	generator	62
ols	5	activator	19	start	60
subsidiary	4	granules	19	variable	60
consigned	3	cycloalkyl	18	codes	52
realm	3	peroxygen	18	interfacing	49
technique	126	activators	17	decoding	47
sufficiently	22	salts	12	coded	39
applicable	84	acetonitrile	11	width	36
wavelength	71	branched	10	extension	35
dramatic	3	diamidodiol	10	mpeg	34
goal	5	methylene	10	tokens	34

Notes: We estimate a latent topic space (i.e., an unknown topic-word distribution) based on the Topic-SCORE approach by Ke and Wang (2024) with three topics. The "Top 20 representative words" are the 20 words with the highest loading for each topic, while "Count" refers to the total number of occurrences of a word across all text documents in the sample. We interpret the latent topics based on the representative words.

The estimated topic-word distribution also provides evidence of significant word frequency heterogeneity in the auction catalogs of patents. Specifically, a fraction of words shows little to no association with any topics, suggesting they serve sentence structure rather than offer meaningful insights. As the majority of the texts directly come from the patent abstract (and claims after treatment), this also explains why examiners and experts play important roles when reviewing patent applications and handling infringement cases. Combined with the aforementioned low degree of semantic similarity among different patents, these textual characteristics highlight the uniqueness of each patent, the challenges in assessing patent values, and the high due diligence

costs faced by potential buyers. Together, these factors underscore the importance of improving information presentation to facilitate IP commercialization. Thus, in the second part of our analysis, we apply a supervised machine learning approach to explore whether topics and representative words can be identified to predict IP commercialization potential.

4. Evaluating Effects of Information Presentation

4.1 Models for Patent Sale Realization

To assess the impacts of better information presentation and the factors influencing patent commercialization, we employ a linear probability model for the binary sale outcome of patent *i* auctioned at year *t* given by:

$$\begin{aligned} Sale_{it} = {} & \beta_1 Post_t + \beta_2 Features_{it} + \beta_3 Citations_{it} + \beta_4 Others_{it} + \beta_5 Words_{it} + \beta_6 Topics_{it} \\ & + \beta_7 Post_t \times Characteristics_{it} + \gamma_1 (\text{Category-fixed effects})_i + \gamma_2 (\text{Auction specific controls})_{it} + e_{it}. \quad (1) \end{aligned}$$

"Post" is a dummy variable, equaling one for patents offered in the auctions after 2007, namely, after recommendations for improving information presentation were received and implemented by the auction house. The other five sets of explanatory variables (named "Features", "Citations", "Others", "Words", and "Topics" as shown in Tables 1 and 2) are included as regressors, except the sellers' estimated value. We take the logarithm of word counts to reduce the influence of extreme outliers with very large values and standardize the level of semantic similarity by its standard deviation across all patent texts, ensuring that the estimated coefficients remain within a reasonable magnitude.

4.2 Empirical Strategies Accounting for Endogeneity

The natural experiment setup renders the exogeneity of the post-treatment dummy. To account for auction-specific effects, we calculate the average characteristics of the patents within

each auction and use these averages as control variables. As a result, as long as the auction house applied the same criteria on observables for selecting and arranging patents in each auction, unobservable time trends and auction-specific factors will not bias the estimation of treatment effects. We also include NAICS industry category fixed effects in the regression because patent values and characteristics such as citation metrics can vary significantly across categories, and different categories may attract different potential buyers. In addition, we examine the sensitivity of the estimated treatment effects to the inclusion or exclusion of the seller's estimated IP value, which is often used to account for unobserved patent and seller heterogeneity. To capture treatment heterogeneity, we interact the treatment variable (Post) with variables of information listed on the patent catalog, as well as the NAICS industry dummies.

There might still be concerns that certain explanatory variables—particularly the seller's estimated IP value, number of listed potential licensees, and the text content (captured by topic weights)—were provided by sellers. Consequently, these variables might be correlated with patent-specific characteristics unobserved to researchers. To address this, we employ the two-stage copula control function estimation (2sCOPE) approach by Yang, Qian, and Xie (2024). The 2sCOPE approach captures the relationship between regressors (both endogenous and exogenous) and the structural error by using a Gaussian copula. This models the rank-based relationships between variables, rather than focusing only on their magnitudes. In other words, it is not just the size of the regressors that matters, but also how their relative rankings influence the outcome, which we exploit to control for endogeneity. It constructs copula-generated regressors to augment the original outcome regression. Conditional on the identification condition of the augmented regression being satisfied – as validated by the data, it allows us to correct endogeneity bias, even in the case with potentially multiple endogenous regressors.

4.3 Exploring Effects on Patent Sale Realization

Table 6 shows the effects of information presentation on whether a patent is sold or not (i.e., realized purchase intent), estimated by OLS and 2sCOPE methods, using a full sample of 1,377 patents after data cleaning in Columns (1)–(5) and a sub-sample of 1,053 patents from those NAICS industries present both before and after the treatment in Column (6).

Both the OLS and 2sCOPE estimates of the treatment effect are significantly positive: patents offered in auctions with improved information presentation are, on average, around 44%–45% more likely to be sold compared to those in the pre-treatment period, confirming that enhanced information presentation facilitates patent transactions. The estimated treatment effect is not affected by whether the seller's estimated value is included as a covariate, or not. Moreover, the endogeneity of certain featured characteristics—those shaped by sellers' strategic decisions—does not appear to bias the estimated total impact of improved information presentation, potentially due to offsetting directions of bias across features.

In contrast to the estimates of the effect of improved information presentation, we find notable discrepancies between OLS and 2sCOPE estimates of the influences of some patent characteristics on the sales models. The 2sCOPE results suggest that selectively featured characteristics, including the seller's estimated value, the number of potential licensees and topic weights, are endogenous, while the variation in other patent characteristics appears to be exogenous.[2]

Extensive research on optimal reserve pricing in auctions highlights trade-offs that similarly apply to how sellers set estimated values in our context. In brief, on one hand, a higher estimated value attracts serious buyers who interpret the estimate as a signal of quality (signaling

[2] For the number of patents in a lot, further analysis reveals that it may be influenced by the nature of the patents themselves, as some patents need to be bundled together to be considered complete and marketable.

Table 6: The Effects of Information Presentation on Patent Sales (0 or 1)

	OLS				2sCOPE						2sCOPE (sub-sample)	
	(1)		(2)		(3)		(4)		(5)		(6)	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Post	0.450	(0.082)	0.450	(0.082)	0.435	(0.216)	0.441	(0.219)	0.335	(0.206)	0.062	(0.557)
1. Seller's estimated IP value												
Log(Estimate)			0.006	(0.011)			-0.059	(0.033)	-0.053	(0.023)	-0.060	(0.041)
2. Features												
No. potential licensees	-0.024	(0.022)	-0.025	(0.022)	0.244	(0.161)	0.233	(0.161)	0.227	(0.093)	0.220	(0.113)
Weight on Topic 2	-0.368	(0.071)	-0.367	(0.071)	-0.795	(0.403)	-0.658	(0.347)	-0.611	(0.277)	-0.709	(0.346)
Weight on Topic 3	0.122	(0.108)	0.122	(0.108)	0.021	(0.456)	-0.038	(0.454)	-0.347	(0.263)	-0.087	(0.574)
3. Citation metrics												
Forward citations	0.004	(0.001)	0.004	(0.001)	0.004	(0.002)	0.003	(0.002)	0.002	(0.001)	0.003	(0.002)
Backward citations	0.0001	(0.0006)	3.E-05	(0.0006)	-0.0001	(0.0005)	0.0002	(0.0005)	0.001	(0.000)	0.001	(0.001)
4. Other Characteristics												
No. patents in a lot	-0.013	(0.006)	-0.013	(0.006)	-0.016	(0.009)	-0.017	(0.009)	-0.015	(0.006)	-0.007	(0.012)
Branded	0.029	(0.030)	0.029	(0.030)	0.013	(0.028)	0.012	(0.027)	0.014	(0.014)	0.001	(0.038)
Patent age	-0.0002	(0.004)	-0.0002	(0.004)	-0.001	(0.004)	-0.002	(0.004)	0.007	(0.004)	-0.001	(0.007)
5. Textual characteristics												
Log(Word count)	0.016	(0.049)	0.011	(0.051)	0.055	(0.055)	0.110	(0.073)	0.052	(0.032)	0.001	(0.060)
Semantic similarity	-0.046	(0.015)	-0.047	(0.015)	-0.072	(0.043)	-0.055	(0.037)	-0.034	(0.020)	-0.052	(0.037)
6. Heterogeneous treatment effects												
Post × Log(Estimate)									0.011	(0.013)	0.025	(0.042)
Post × Forward citations									0.0005	(0.001)	3.E-05	(0.003)
Post × Backward citations									-0.005	(0.002)	-0.005	(0.003)
Post × No. patents in a lot									0.034	(0.015)	0.036	(0.014)
Post × Patent age									-0.026	(0.012)	-0.021	(0.012)
7. Auction-specific effects												
Max Log(Estimate)	0.343	(0.074)	0.339	(0.074)	0.310	(0.161)	0.371	(0.187)	0.308	(0.133)	0.248	(0.092)
Avg. No. potential licensee	0.138	(0.041)	0.141	(0.041)	-0.125	(0.141)	-0.122	(0.142)	-0.082	(0.059)	-0.077	(0.124)
Avg. forward citations	-0.026	(0.006)	-0.026	(0.006)	-0.024	(0.013)	-0.025	(0.013)	-0.025	(0.011)	-0.021	(0.007)
Avg. backward citations	0.035	(0.009)	0.035	(0.009)	0.038	(0.021)	0.038	(0.021)	0.038	(0.016)	0.035	(0.017)
Avg. patent age	-0.094	(0.023)	-0.096	(0.023)	-0.092	(0.048)	-0.078	(0.042)	-0.063	(0.030)	-0.078	(0.033)
Correlations with estimated structural errors												
Log(Estimate)							0.067	(0.044)	0.065	(0.030)	0.085	(0.059)
No. potential licensees					-0.224	(0.129)	-0.205	(0.125)	-0.189	(0.075)	-0.218	(0.087)
Weight on Topic 2					0.085	(0.055)	0.056	(0.048)	0.024	(0.022)	0.060	(0.099)
Weight on Topic 3					-0.066	(0.105)	-0.036	(0.102)	0.035	(0.051)	-0.011	(0.122)
No. observations	1,377		1,377		1,377		1,377		1,377		1,053	

Notes: The effects of information presentation on patent sales are estimated based on 1,377 patents in eight auctions with patent NAICS category fixed effects. "Post" is a dummy variable, equaling one for patents offered in auctions after 2007. For other explanatory variables, see Tables 2, 3, and 4. Numbers in brackets are estimated standard errors, where bootstrap is used for estimated indirect and total effects and all 2sCOPE estimates.

effect) and may anchor bids upward, leading to a higher final price (anchoring effect). On the other hand, setting a higher estimated value may discourage potential buyers with liquidity constraints from participating (screening or deterrence effect). When sellers are required to invest greater effort in structuring and presenting information to potential buyers, they are prompted to weigh the benefits of drawing in more interest through lower estimates against the risk of receiving lower bids. As a result, sellers tend to be more conservative in setting their estimated market values, recognizing that in an environment of uncertainty—where buyers must conduct their own due diligence and are less willing to commit large sums—overstated estimates may ultimately reduce buyer engagement.[3] The resulted sellers' estimates became more closely aligned with reserve values rather than aspirational transaction prices.

Comparing Columns (2) and (4), accounting for the endogeneity of Log(Estimate) notably affects the estimate of its effect on IP sales. The 2sCOPE estimation result yields a positive correlation between the estimated value and the structural error at 0.07 (se. = 0.04), suggesting that sellers' estimated values are signals for patents unobserved quality.[4] Nevertheless, the deterrence effect of sellers' estimated values tends to dominate, explaining the significant negative impact of the estimated value on sale outcomes around –0.06. Namely, a one-unit increase in the log of the estimated value—raising it from the mean of 1,303,951 (averaged across all 1,377 patents) to approximately 2,248,200—reduces the probability of sale by 6%.[5] In contrast, the OLS estimates show a negligible effect of Log(Estimate).

[3] Despite that the OLS and 2sCOPE estimates of the treatment effect on Log (Estimate) are close to each other, we find substantial differences in the coefficients of the selectively featured patent characteristics, which suggests the use of 2sCOPE for understanding how patent characteristics affect sellers' estimated market values.
[4] Figures W1–W3 in the web appendix provide evidence that the empirical distributions of Log (Estimate), number of potential licensees, and weights on topic 2 and topic 3 are non-normal distributed such that the 2sCOPE method has identification powers.
[5] Notice that 2,248,200 is close to the 75th percentile of estimated values pre-treatment (2,000,000) but is significantly higher than the 75th percentile post-treatment (550,000).

Moreover, the 2sCOPE estimates reveal a negative correlation between the number of potential licensees at -0.224 (se. = 0.129) and the unobserved endogenous error term. The corresponding correlation coefficients for the weights on topic 2 and topic 3 are 0.085 (se. = 0.002) and -0.067 (se. = 0.014), respectively. This suggests that the structural error contains unobserved patent characteristics valued by sellers but not fully recognized by buyers, as well as these unobserved factors influencing sellers' decisions about which IPs to offer at auction, which we expect to be accounted for by the 2sCOPE method. Specifically, sellers may overstate their IP's commercialization potential by listing outreach to potential licensees, which—if perceived as exaggerated by buyers—can reduce purchase intent. The resulting downward bias, combined with the effect's statistical insignificance both before and after bias correction, suggests that this measure serves as a noisy signal. We also find that the negative impact of the weight on topic 2, 'Chemicals,' on patent sales is substantially underestimated. This may reflect sellers' awareness of limited demand for IPs in this category and their decision to offer these patents at auction recognizing the difficulty of reaching potential buyers through other channels.

Comparing Columns (3) and (4), including the seller's estimated price as a covariate primarily changes the estimated effects of textual features. Since sellers base their price expectations on the content and sentiment of the text, their estimated values likely capture part of the influence of IP textual characteristics on outcomes. As shown in column (4), after including the seller's estimated price as a covariate, we find no significant impact of word count or semantic similarity but topic weights. With Topic 1 as the reference group, the negative coefficients on Topics 2 and 3 indicate that patents more closely associated with Topic 1 are more likely to sell. Topic 2, representing a distinct technology class, appears to be less valued or attracts fewer buyers, as indicated by its negative effect. While Topics 1 and 3 are more closely related, Topic 1 features

more attention-grabbing adjectives, whereas Topic 3 is characterized by functional nouns, which may reduce engagement and divert buyer attention. These results also highlight the informativeness of machine learning-derived textual features in shaping sellers' and buyers' valuation. Now we focus on the 2sCOPE results in Column (4) and discuss effects of other patent characteristics (value drivers) on patent sales. In terms of citation metrics, we consistently find a positive effect in having more forward citations but an insignificant effect of the backward citations. We observe a negative effect of the number of patents bundled in a lot, likely reflecting diminished buyer interest due to budget constraints.[6]

Next, we examine auction-specific control variables, which are calculated as the average characteristics of patents within each auction, except Log (Estimate), where we use the maximum value across patents. Interestingly, we find a consistently significant positive effect of the maximum seller's estimated IP value on patent sales. Specifically, a roughly 2.7-fold increase in the highest estimated value within an auction is associated with a 37.1% increase in the likelihood of sale, suggesting the presence of another type of anchoring effects driven by within-auction comparisons. The impact of other auction-specific controls computed as averages can be interpreted as indicating complementary or substitution effects among patents offered in the same auction. In general, a patent achieves a better sales outcome in an auction where the top-valued patent has a greater estimated market value and, *on average*, the patents being offered in the same auction have fewer forward citations (less "general"), more backward citations (less "original"), and later expiration dates. While the three factors may render an intensive comparative advantage to a patent, a lower estimated value on average may help attract more bidders due to affordability, thus providing benefits to a patent on an extensive margin.

[6] The squared term of the number of patents is included to account for potential nonlinear effects.

Column (5) presents heterogeneous treatment effects of information presentation on patent sales, estimated using the 2sCOPE method with interaction terms between the treatment and non-textual patent characteristics. The results reveal significant moderation effects: improved information presentation is particularly effective for patents with fewer backward citations, larger lot sizes, and later expiration dates. These characteristics amplify the positive impact of presentation quality, highlighting their role in shaping how buyers interpret and respond to disclosed information. The above results remain robust when we conduct the analysis using a sub-sample that includes only 1,053 patents from those NAICS industries present both before and after the treatment, as shown in Column (6).

4.4 Analyses of Patent Lot Sale Price

Since the prices of unsold patent lots are unobserved, to assess the impact of improved presentation, we use a sub-sample of sold patent lots, comprising 468 patents—305 before treatment and 163 after treatment.[7] We consider a linear model for the logarithm of the final price of patent lot *j* auctioned at year *t* given by:

$$\begin{aligned}\text{Log}(\text{Price}_{jt}) = \beta_1\text{Post}_t + \beta_2\text{Features}_{jt} + \beta_3\text{Citations}_{jt} + \beta_4\text{Others}_{jt} + \beta_5\text{Words}_{jt} + \beta_6\text{Topics}_{jt} \\ + \beta_7\,\text{Post}_t \times \text{Characteristics}_{jt} + \gamma_1\,(\text{Category-fixed effects})_j + \gamma_2\,(\text{Auction specific controls})_{jt} + u_{jt}. \quad (4)\end{aligned}$$

Table 7 summarizes the effects of information presentation on Log(Price) of sold patents in auctions. The OLS estimated treatment effect has a substantial upward bias due to contamination

[7] As can be seen in the left sub-plot of Figure 2, a larger fraction of patents remained unsold in the pre-treatment period, as indicated by the higher density near zero. Additionally, post-treatment patents were sold at relatively higher prices, as seen from the greater density in the right tail of the distribution. Considering these patterns, including both sold and unsold patents in the analysis of sale prices could overstate the estimated treatment effect due to many zeroes in the data set. It may also fail to capture the distinct distribution patterns of positive sale prices, which exhibit a bimodal structure, as shown in the right sub-plot of Figure 2. Moreover, since "Estimate" reflects the seller's expected sale or market price—which can exceed the reserve price, as observed in the data—we cannot use Log(Estimate) as a proxy for Log(Price) for unsold patents.

Table 7: The Effects of Information Presentation on Log(Price) of Sold Patents

	OLS		2sCOPE		2sCOPE with heterogeneity		2sCOPE in sub-sample	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Post	1.152	(0.176)	0.544	(0.322)	-0.810	(0.506)	0.003	(1.431)
1. Features								
Log(Estimate)	0.638	(0.032)	0.589	(0.245)	0.628	(0.252)	0.746	(0.099)
No. potential licensees	0.128	(0.044)	0.337	(0.162)	0.216	(0.140)	0.037	(0.277)
Weight on Topic 2	-0.606	(0.347)	-1.421	(0.580)	-1.198	(0.545)	0.085	(1.649)
Weight on Topic 3	-0.483	(0.254)	-4.263	(1.811)	-2.669	(1.080)	-2.005	(1.759)
2. Textual characteristics								
Log(Word count)	-0.664	(0.164)	-1.036	(0.446)	-0.886	(0.379)	-0.835	(0.276)
Semantic similarity	0.049	(0.043)	0.099	(0.073)	0.045	(0.059)	0.070	(0.125)
3. Citation metrics								
Forward citations	-0.0002	(0.002)	-0.003	(0.001)	-0.0003	(0.002)	-0.002	(0.006)
Backward citations	-0.005	(0.002)	-0.005	(0.003)	-0.007	(0.003)	-0.008	(0.004)
4. Other Characteristics								
No. patents in a lot	0.203	(0.036)	0.230	(0.091)	0.203	(0.081)	0.090	(0.066)
Squared no. patents in a lot	-0.010	(0.002)	-0.011	(0.004)	-0.010	(0.004)	-0.005	(0.003)
Branded	0.010	(0.069)	0.115	(0.079)	0.060	(0.055)	0.108	(0.134)
Patent age	0.020	(0.010)	0.059	(0.026)	0.041	(0.017)	0.042	(0.029)
5. Heterogeneous treatment effects								
Post × Log(Estimate)					0.143	(0.059)	0.036	(0.110)
Post × No. potential licensees					-0.170	(0.114)	0.010	(0.290)
Post × Forward citations					-0.002	(0.003)	0.002	(0.007)
Post × Backward citations					0.005	(0.003)	0.006	(0.005)
Post × No. patents in a lot					0.022	(0.017)	0.069	(0.028)
Post × Patent age					0.008	(0.009)	0.000	(0.023)
6. Auction-specific effects								
Max Log(Estimate)	-0.305	(0.144)	-0.647	(0.302)	-0.654	(0.273)	-0.614	(0.250)
Average no. potential licensees	0.077	(0.087)	-0.333	(0.148)	-0.063	(0.096)	0.099	(0.259)
Average forward citations	0.049	(0.011)	0.086	(0.035)	0.062	(0.025)	0.063	(0.021)
Average backward citations	0.062	(0.025)	-0.077	(0.048)	-0.017	(0.029)	-0.030	(0.080)
Average patent age	-0.789	(0.065)	-0.668	(0.279)	-0.686	(0.285)	-0.784	(0.107)
Correlations with estimated structural errors								
Log(Estimate)			0.068	(0.037)	-0.013	(0.032)	-0.044	(0.090)
No. potential licensees			-0.063	(0.054)	0.006	(0.049)	0.084	(0.127)
Weight on Topic 2			-0.136	(0.073)	-0.084	(0.065)	-0.235	(0.154)
Weight on Topic 3			0.454	(0.175)	0.313	(0.134)	0.352	(0.202)
Number of observations	468		468		468		405	

Notes: The effects of information presentation on patent sales are estimated using sold patents from eight auctions. All models include patent NAICS category fixed effects, and models with heterogeneous treatment effects, in addition, include interaction terms between the "Post" indicator and category dummies. "Post" is a dummy variable, equaling one for patents offered in the auctions after 2007. For the other explanatory variables, see Tables 2, 3, and 4. Numbers in brackets are estimated standard errors, where bootstrap is used for estimated indirect and total effects and all 2sCOPE estimates.

from the endogeneity of selectively featured patent characteristics. Based on the 2sCOPE estimation results, the main effect of improved information presentation indicates that the average

price increases by approximately 72.3% in the post-treatment period. Moreover, among patents that were successfully sold, the seller's estimated value has a significantly positive effect on the final sale price, which is consistent with findings in Roberts (2013). Specifically, a 10% increase in the estimated value corresponds to a 6% increase in the actual sale price, on average, holding all other factors constant. This also confirms the trade-off faced by sellers when setting their targeted or estimated patent values—higher estimates may lead to higher prices but also reduce the likelihood of a sale. Furthermore, improving information presentation strengthens the positive relationship between estimated values and patent sale prices. Specifically, the coefficient of Post × Log(Estimate) indicates an additional 0.14% increase in the sale price effect, on top of the 0.6% baseline coefficient for Log(Estimate).

When examining other drivers of patent value, we find that backward citations have a significant negative effect, suggesting that buyers place higher value on more original patents—those with fewer references. However, if buyers are unwilling to meet the seller's reserve price, for example, due to uncertainty in its valuation, these more original patents may be less likely to sell successfully, as shown in Table 6. This effect on price is reduced by the improved information presentation as shown by the positive coefficients in the interaction term with "Post". The number of patents in a lot has a positive effect on the total sale price, as expected. However, the magnitude of this effect suggests that the sale price does not increase proportionally with the number of patents pooled in a lot; instead, it grows at a much slower rate. Also, older patents are associated with higher prices.

The influences of textual characteristics are much stronger on sale prices. Specifically, they have already incurred higher due diligence costs for inspecting these patents and are therefore less willing to pay a higher transaction price. Also, patent catalog descriptions that are comprised more

of Topics 2 and 3 are also associated with lower prices. These findings motivate us to examine the impact of narratives using a supervised machine learning approach in the following sections.

Last but not the least, we find that the final prices of sold patents decrease when the maximum estimated IP value within an auction is higher. This may be because buyers are drawn to the most highly valued item and allocate less of their budget to other patents in the same auction. As for auction-specific factors, auctions with a higher average number of forward citations per patent and a greater share of younger patents tend to achieve higher sale prices. However, while more forward citations per patent within an auction can raise final bid prices of sold items, they may also lead to fewer patents being sold within an auction due to buyers' budget constraints, as shown in Table 6.

5. A Supervised Learning Approach for Patent Sale Prediction

Using the estimated topic weights from the unsupervised Topic-SCORE approach, we identified significant impacts on sales outcomes and observed a shift in topics across auctions before and after the treatment. In this section, we examine which specific features of patent text documents are key factors influencing the prediction of patents' commercialization potential. Unlike unsupervised methods, we adopt the heterogeneous supervised topic model (HSTM) introduced by Sridhar, Daumé III, and Blei (2022), which functions as both a generative model for text and a supervised model for classification.

There are three main advantages of using HSTM: first, the HSTM model incorporates effects of both topics and words (or tokens) on the outcome prediction; second, the effects are allowed to be heterogeneous for different documents; and third, in addition to the representative ("neutral") words for each topic, the model can uncover words associated with the strongest positive ("pro") or negative ("passive") tones towards the sale probability within a topic and across

text documents. Thus, these uncovered latent topics as well as topic-specific and overall pro-sale words, can be used to search for and identify patents with a higher likelihood of being sold for the auction house (or other platforms) to host future sales events.[8]

Table 8 summarizes the performance of HSTM with different topic numbers across 50 different random seeds used in the Bayesian estimation process. The model's ability to classify sold and unsold lots is assessed using the area under the Receiver Operating Characteristic curve (ROC AUC). The range of the accuracy scores is comparable to those reported by Suzgun et al. (2023), which compares various large language models for the binary classification of patent acceptance.

Table 8: Performances of the HSTM Model on Predicting the Binary Patent Sale Outcomes

	Receiver Operating Characteristic Area Under the Curve			
No. topics	Mean	SD	Median	Max
5	0.55	0.05	0.54	0.65
10	0.55	0.05	0.55	0.64
15	0.56	0.04	0.56	0.67
20	0.55	0.04	0.55	0.68
25	0.56	0.03	0.55	0.62
30	0.55	0.03	0.54	0.62
35	0.56	0.03	0.56	0.66
40	0.55	0.03	0.55	0.64
45	0.55	0.03	0.56	0.62
50	0.56	0.03	0.56	0.63
55	0.55	0.03	0.55	0.59
60	0.56	0.02	0.56	0.60

Notes: This table shows performances of the HSTM model on the binary classification of patent sale outcomes in test sets, averaged over 50 random seeds that determine. The Receiver Operating Characteristic Area Under the Curve (ROC AUC) measures the area under the ROC curve, which shows the trade-off between true positive rate and false positive rate across thresholds, indicating the model's ability to distinguish between two classes with a score from 0 to 1.

The HSTM model achieves its best classification performance with 20 topics, reaching a peak ROC AUC of 0.68. Beyond 20 topics, further increases offer minimal gains in classification

[8] The HSTM model by Sridhar, Daumé III, and Blei (2022), nonetheless, does not integrate other structured covariates into the prediction, which is an area for future research.

accuracy. Intuitively, due to the high uncertainty in patent commercial values, predicting patent sale outcomes requires more adaptable word clusters ("topics"). It comes at the cost of higher uncertainty of generated texts when the underlying latent topic structure is relatively low-dimensional. The improvements in predictive power will eventually plateau and decline when additional topics become redundant.

Table 9 provides a visualization of the overall pro-sale words and passive words identified by HSTM models with 15 and 20 topics, based on the seeds that yielded the highest ROC AUC score above or equal to 0.67. First, we find that pro-words and passive words belong to distinct and refined technology fields. Pro-words are linked to technologies such as "computer-based" systems, "information servers," "interactive television" and "transport." For instance, terms like "control circuit" and "converter" may refer to hardware components of an information server, while phrases such as "decoding," "information stored," and "message transmitted" describe the functional aspects of patented technologies. In contrast, passive words are associated with technologies related to "manufacturing," "materials" and "cell phone." Second, the verbs associated with pro-words and passive words differ in tone. Pro-words, such as "deals," "maximize," "transform," "manipulating," and "implemented," generally convey a more assertive and action-oriented tone, while passive words like "accounting," "impact," and "improves," tend to have a softer and more neutral tone. Lastly, certain words also reflect buyers' preferences. Words related to effectiveness, such as "improvements" and "impacts," tend to align more with incremental advancements rather than breakthrough innovations and are found among the top passive words.

Although the words presented in Tables 5 and 9 differ, this is expected, as the unsupervised and supervised topic models we employ are designed to uncover distinct latent patterns in patent

Table 9: Top 65 Overall Pro-Words and Passive Words Identified by HSTM Models

Overall pro:	amplifiers, audio signal, cacheable, channels, characters, client, columns, computer based, control circuit, converter, couplings, data second, deals, decoded, decoding, deleting, describing, elementary, enhancement, first second, formats, frames, graphical user, handsets, html, implemented, included lot, information including, information least, information regarding, information server, information stored, interactive television, location data, locations, lost stolen, manipulating, maximize, message transmitted, metadata, motion, movies, multiple access, one plurality, output, opportunities, patent offered, programmer, programs, rejected, set, spray, stolen, stored said, subset, streaming, system provides, transform, transport, vehicular, via, video, videos, viewer, wheel.
Overall passive:	accounting, acid, active, article, aspect, attachment, bar code, cell, cell phone, comprises, connect, contacting, delivery, designs, diameter, diffraction, disclosure, displays, electrode, flexibility, flow, fluid flow, formula, funds, gas, handheld device, high frequency, impact, improves, japan, lack, lead, manufacture, manufacturing, materials, membrane, method manufacturing, noninvasive, opening, oxide, paper, porous, portion, primary, print, produce, purchased, receiving party, requirements, reserve, rf, rf signals, silicone, subscribers, support, surface, take, technology disclosed, telephone number, therein, time day, titanium, transparent, without need.

Notes: The top 65 pro- and passive words are consistently identified among the top 200 pro- and passive words in HSTM models with 15 and 20 topics, based on the seeds that achieved the highest ROC AUC scores above or equal to 0.67.

texts, each serving a different purpose. Unsupervised models, such as Topic-SCORE, excel at identifying a smaller set of well-defined and cohesive topics, while supervised models, like HSTM, are more effective for prediction-oriented tasks and extracting outcome-driven insights. These findings highlight how leveraging a supervised topic model allows us to learn from past auctions and inform future decision-making.

6. Discussions and Conclusions

In this section, we begin by summarizing our findings from the natural experiment designed to address our research questions, followed by a discussion of their implications for IP managers.

First, does better information presentation facilitate IP monetization? Using IP data from

eight consecutive auctions, we find that information presentation significantly impacts IP transactions. Through auctions, Ocean Tomo has reduced the marketing costs for sellers. However, to bridge the gap between seller and buyer valuations in the presence of significant information asymmetry and uncertain future payoffs, our analysis shows that sellers and auction houses should present information in a clear, organized format with comprehensible text summaries. These procedures can help attract more buyers and encourage them to invest in further evaluation, such as conducting due diligence which is costly but inevitable in the market for intellectual property. From a marketing perspective, this finding highlights how effective information design can reduce perceived risk and improve buyer engagement, especially in high-uncertainty, low-comparability contexts.

Second, do IP characteristics moderate the impact of better information presentation? Our findings demonstrate that the effect of information presentation on patent value is not uniform and is moderated by underlying IP characteristics. Specifically, improved presentation has a greater impact for lots that are more complex (i.e., include more bundled patents), more original (i.e., have fewer backward citations), and have later expiration dates, indicating that well-structured information is particularly valuable for transactions of harder-to-evaluate IP. What's more, our empirical findings show that while higher seller-estimated values may reduce the likelihood of a sale, they tend to lead to higher final prices when a sale does occur. Importantly, enhanced information presentation strengthens the positive impact of these estimated values on sale prices. Together, these results suggest that the value buyers assign to disclosed information is shaped by the nature of the IP itself, underscoring the importance of tailoring presentation strategies to different types of patents. This aligns with theories of cognitive fit, which argue that the presentation format should match the structure of the decision task to improve effectiveness.

Furthermore, our prediction exercise mirrors the patent examination process, where examiners first classify a patent application into one or more technology categories and then search for prior art to evaluate its novelty. Similarly, using a supervised learning approach, we categorize words into topics and identify pro-words and passive words within each topic and across all documents to infer the sale probabilities of unseen patents. We noticed that many patent texts in auction catalogs are directly copied from patent abstracts without modification. Given that our models reveal the heterogeneous impacts of different words on predicting sale probability, it will be beneficial to draft summaries that are both clearer and more engaging while maintaining accurate descriptions of the patent's functionality. Such findings provide actionable implications for the marketing content of intangible assets.

In a web appendix, we document a supporting laboratory experiment designed to replicate our field findings and explore the underlying mechanisms of why information presentation affects buyer perception, especially for products such as IP, whose features are not easily searchable or observable at the time of purchase. The experiment shows that salient and organized information leads to higher perceived quality, more favorable attitudes, and greater perceived value. These results reinforce our field-based conclusions and offer actionable implications for how marketing content should be structured when communicating intangible assets.

Our study has important implications for both IP valuation and auction practices. We present strong evidence that information presentation affects IP value. Our findings offer guidance on which patent characteristics should be highlighted in IP marketing and how these features impact value, contributing to the IP valuation literature. Our study also provides unique empirical tests and extensions to the theory of cognitive fit (Vessey, 1991) and the linkage principle in auctions (Milgrom and Weber, 1982), by revealing that it is not only the quantity and type of

information released that affects auction price, but also the presentation of information, in terms of its understandability and usefulness, matters. We theoretically generalize these findings in the Appendix. This theoretical extension has general implications for information disclosure and auction design beyond IP settings.

This research has several direct managerial implications for IP professionals. First, the findings indicate that IP information prospectus offerings should be carefully crafted. The information about the IP should be organized in a way that makes the important value drivers (i.e., IP characteristics) easy to understand. Second, this research has implications for IP sellers as it addresses issues related to IP marketing during other transaction situations (e.g., mergers and acquisitions). For example, presenting and organizing IP into groups of patents that reflect particular technology areas, commercial importance, and key value drivers will help sellers generate higher values for their IP. Third, IP buyers benefit from this research as it identifies important drivers of IP value that are not always obvious, even to experienced IP professionals. Fourth, this research documents and analyzes an emerging form of marketing and commercialization of IPs. The advantages of an auction format, compared to traditional licensing and private transaction, rest in the marketing tactic of bringing together multiple potential buyers to create a sense of demand and urgency in making decisions. This further calls for effective and efficient presentation of information to facilitate timely transactions. Finally, this research has direct implications for auction house managers as it illustrates the importance and value of providing marketing information in an organized format to generate better performance.

Despite the unique setting of our natural experiment on patent auctions hosted by Ocean Tomo, we recognize limitations, particularly regarding data access from other patent platforms, which could affect the generalizability of our findings, for example, to other technology classes.

Nonetheless, this study opens a new research avenue in IP marketing. While prior research has largely focused on the legal, competitive, and strategic dimensions of intangible assets, our study highlights the role of marketing, specifically information presentation, in enhancing IP commercial value. This critical element deserves serious attention from IP managers aiming to optimize IP portfolios.

References

Anton, James J., and Dennis A. Yao (2002), "The sale of ideas: Strategic disclosure, property rights, and contracting," *The Review of Economic Studies*, 69 (3), 513-531.

Anton, James J., and Dennis A. Yao (2008), "Attracting skeptical buyers: Negotiating for intellectual property rights," *International Economic Review*, 49 (1), 319-348.

Arora, Ashish, and Alfonso Gambardella (2010), "Ideas for rent: An overview of markets for technology," *Industrial and Corporate Change*, 19 (3), 775-803.

Arora, Ashish, Sharon Belenzon, and Jungkyu Suh (2022), "Science and the market for technology," *Management Science*, 68 (10), 7176-7201.

Arts, Sam, Jianan Hou, and Juan Carlos Gomez (2021), "Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures," *Research Policy*, 50 (2), 104144.

Arts, Sam, Bruno Cassiman, and Jianan Hou (2023), "Position and differentiation of firms in technology space," *Management Science*, 69 (12), 7253-7265.

Aubry, Mathieu, Roman Kräussl, Gustavo Manso, and Christophe Spaenjers (2023), "Biased auctioneers," *The Journal of Finance*, 78 (2), 795-833.

Bessen, James (2008), "The value of US patents by owner and patent characteristics," *Research Policy*, 37 (5), 932-45.

Bettman, James R., and Pradeep Kakkar (1977), "Effects of information presentation format on consumer information acquisition strategies," *Journal of Consumer Research*, 3 (4), 233-240.

Beyhaghi, Mehdi, Pooyan Khashabi, and Ali Mohammadi (2023), "Pre-grant patent disclosure and analyst forecast accuracy," *Management Science*, 69 (5), 3140-3155.

Biehal, George, and Dipankar Chakravarti (1982), "Information-presentation format and learning goals as determinants of consumers' memory retrieval and choice processes," *Journal of Consumer Research*, 8 (4), 431-441.

Bollinger, Bryan, Phillip Leslie, and Alan T. Sorensen (2011), "Calorie posting in chain restaurants". *American Economic Journal: Economic Policy*, 3(1), 91–128.

Bronnenberg, Bart J., Jean-Pierre Dubé, and Matthew Gentzkow (2012), "The evolution of brand preferences: Evidence from consumer migration". *American Economic Review*, 102(6), 2472–2508.

Caviggioli, Federico, and Elisa Ughetto (2016), "Buyers in the patent auction market: Opening the black box of patent acquisitions by non-practicing entities," *Technological Forecasting and Social Change*, 104, 122-132.

Caviggioli, Federico, Antonio De Marco, Giuseppe Scellato, and Elisa Ughetto (2017), "Corporate strategies for technology acquisition: Evidence from patent transactions," *Management Decision*, 55 (6), 1163-1181.

Chen, Mark A., Qinxi Wu, and Baozhong Yang (2019), "How valuable is FinTech innovation?" *The Review of Financial Studies*, 32 (5), 2062-2106.

Chesbrough, Henry (2006), "Emerging secondary markets for intellectual property: US and Japan comparisons," *National Center for Industrial Property Information and Training*.

Choi, Jaewoong, Byeongki Jeong, Janghyeok Yoon, Byoung-Youl Coh, and Jae-Min Lee (2020), "A novel approach to evaluating the business potential of intellectual properties: A machine learning-based predictive analysis of patent lifetime," *Computers & Industrial Engineering*, 145, 106544.

Cohen, Jacob (1960), "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, 20 (1), 37-46.

De Marco, Antonio, Giuseppe Scellato, Elisa Ughetto, and Federico Caviggioli (2017), "Global markets for technology: Evidence from patent transactions," *Research Policy*, 46 (9), 1644-1654.

Dushnitsky, Gary, and Thomas Klueter (2017), "Which industries are served by online marketplaces for technology?," *Research Policy*, 46 (3), 651-666.

Einhorn, Hillel J. and Robin M. Hogarth (1981), "Behavior decision theory: Processes of judgment and choice," *Annual Review of Psychology*, 32, 53-88.

Fan, Cuihong, Byoung Heon Jun, and Elmar G. Wolfstetter (2016), "Optimal bid disclosure in patent license auctions under alternative modes of competition." International Journal of Industrial Organization, 47, 1-32.

Figueroa, Nicolás, and Carlos J. Serrano (2019), "Patent trading flows of small and large firms," *Research Policy*, 48 (7), 1601-1616.

Fischer, Timo, and Jan Leidinger (2014), "Testing patent value indicators on directly observed patent value—An empirical analysis of Ocean Tomo patent auctions," *Research Policy*, 43 (3), 519-529.

Fradkin, Andrey, and David Holtz (2023). Do incentives to review help the market? Evidence from a field experiment on Airbnb. *Marketing Science*, 42(5), 853-865.

Furman, Jeffrey L., Markus Nagler, and Martin Watzinger (2021), "Disclosure and subsequent innovation: Evidence from the patent depository library program." *American Economic Journal: Economic Policy*, 13 (4), 239-270.

Galasso, Alberto, Mark Schankerman, and Carlos J. Serrano (2013), "Trading and enforcing patent rights," *The RAND Journal of Economics*, 44 (2), 275-312.

Glazer, Rashi (1991), "Marketing in an information-intensive environment: Strategic implications of knowledge as an asset," *Journal of Marketing*, 55 (4), 1-19.

Gambardella, Alfonso, Paola Giuri, and Alessandra Luzzi (2007), "The market for patents in Europe," *Research Policy*, 36 (8), 1163-1183.

Gambardella, Alfonso, Dietmar Harhoff, and Bart Verspagen (2008), "The value of European patents," *European Management Review*, 5 (2), 69-84.

Gans, Joshua S. and Scott Stern (2010), "Is there a market for ideas?" *Industrial and Corporate Change*, 19 (3), 805-37.

Godes, David and Dina Mayzlin (2004), "Using online conversations to study word-of-mouth communication." *Marketing Science*, 23(4), 545–560.

Hagiu, Andrei, and David B. Yoffie (2013), "The new patent intermediaries: Platforms, defensive aggregators, and super-aggregators," *Journal of Economic Perspectives*, 27 (1), 45-66.

Harbert, Tam (2015), "Google versus the trolls," *IEEE Spectrum*, 52 (11), 14-14.

Harhoff, Dietmar, Frederic M. Scherer, and Katrin Vopel (2003), "Citations, family size, opposition and the value of patent rights," *Research Policy*, 32 (8), 1343-63.

Hegde, Deepak, and Hong Luo (2018), "Patent publication and the market for ideas," *Management Science*, 64 (2), 652-672.

Hong, Xianpei, Menghuan Zhou, and Yeming Gong (2021), "Dilemma of quality information disclosure in technology licensing," *European Journal of Operational Research*, 294 (2), 543-557.

Jarosz, John (2010), "Patent auctions: How far have we come?" *les Nouvelles* (March 2010).

Jarvenpaa, Sirkka L. (1989), "The effect of task demands and graphical format on information processing strategies," *Management Science*, 35 (3), 285-303.

Jeon, Haejun (2019), "Licensing and information disclosure under asymmetric information," *European Journal of Operational Research*, 276 (1), 314-330.

Jin, Ginger Zhe, and Andrew Kato (2006), "Price, quality, and reputation: Evidence from an online field experiment." *RAND Journal of Economics,* 37(4), 983–1005.

Kagel, John H. and Dan Levin (1986), "The winner's curse and public information common value auctions," *The American Economic Review*, 76 (5), 894-920.

Kang, Byeongwoo, Rudi Bekkers, and Atsushi Ohyama (2024), "The determinants of parallel invention: Measuring the role of information sharing and personal interaction between inventors," *Research Policy*, 53 (10), 105092.

Ke, Zheng Tracy, and Minzhe Wang (2024), "Using SVD for topic modeling," *Journal of the American Statistical Association*, 119 (545), 434-449.

Kelton, Andrea S., Robin R. Pennington, and Brand M. Tuttle (2010), "The effects of information format on judgment and decision making: A review of the information systems research," *Journal of Information Systems*, 24 (2), 79-105.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman (2017), "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, 132 (2), 665-712.

Kwon, Jung H., Haemin Dennis Park, and Shu Deng (2022), "When do firms trade patents?," *Organization Science*, 33 (3), 1212-1231.

Lacetera, Nicola, Bradley J. Larsen, Devin G. Pope, and Justin R. Sydnor (2016), "Bid takers or market makers? The effect of auctioneers on auction outcome," *American Economic Journal: Microeconomics*, 8 (4), 195-229.

Lück, Sonja, Benjamin Balsmeier, Florian Seliger, and Lee Fleming (2020), "Early disclosure of invention and reduced duplication: An empirical test," *Management Science*, 66 (6), 2677-2685.

Malackowski, James E. (2006), "The intellectual property marketplace: Past, present and future," *The John Marshall Review of Intellectual Property Law*, 5, 605-616.

Malhotra, Naresh K. (1982), "Information load and consumer decision making," *Journal of Consumer Research*, 8 (4), 419-430.

Milgrom, Paul R. and Robert J. Weber (1982), "A Theory of Auctions and Competitive Bidding," *Econometrica: Journal of the Econometric Society*, 50 (5), 1089-1122.

Milgrom, Paul R. (2004). *Putting Auction Theory to Work*. Cambridge University Press.

Odasso, Cristina, Giuseppe Scellato, and Elisa Ughetto (2015), "Selling patents at auction: An empirical analysis of patent value," *Industrial and Corporate Change*, 24 (2), 417-438.

Reuer, Jeffrey J., Sandip Bisui, and George Chondrakis (2025), "Innovation disclosures and the design of technology acquisition contracts: Evidence from the American Inventors Protection Act," *Strategy Science*, 10 (1), 68-92.

Roberts, James W (2013), "Unobserved heterogeneity and reserve prices in auctions." *The RAND Journal of Economics*, 44 (4), 712-732.

Seiler, Stephan (2013), "The impact of search costs on consumer behavior: A dynamic approach". *Quantitative Marketing and Economics*, 11(2), 155–203.

Serrano, Carlos J (2010), "The dynamics of the transfer and renewal of patents," *The RAND Journal of Economics*, 41 (4), 686-708.

Serrano, Carlos J (2018), "Estimating the gains from trade in the market for patent rights," *International Economic Review*, 59 (4), 1877-1904.

Simeth, Markus, and Michele Cincera (2016), "Corporate science, innovation, and firm value." *Management Science,* 62 (7), 1970-1981.

Sneed, Katherine A. and Daniel K. N. Johnson (2009), "Selling ideas: The determinants of patent value in an auction environment," *R&D Management*, 39 (1), 87-94.

Sood, Ashish, and Gerard J. Tellis (2009), "Do innovations really pay off? Total stock market returns to innovation," *Marketing Science*, 28 (3), 442-456.

Sridhar, Dhanya, Hal Daumé III, and David Blei (2022), "Heterogeneous supervised topic models," *Transactions of the Association for Computational Linguistics*, 10, 732-745.

Suzgun, Mirac, Luke Melas-Kyriazi, Suproteem Sarkar, Scott D. Kominers, and Stuart Shieber (2023), "The Harvard USPTO patent dataset: A large-scale, well-structured, and multi-purpose corpus of patent applications." *Advances in Neural Information Processing Systems*, 36, 57908-57946.

Tadelis, Steven and Florian Zettlemeyer (2015), "Information disclosure as a matching mechanism: Theory and evidence from a field experiment," *American Economic Review*, 105 (2), 886-905.

Testoni, Marco (2022), "The market value spillovers of technological acquisitions: Evidence from patent-text analysis," *Strategic Management Journal*, 43 (5), 964-985.

Todd, Peter and Izak Benbasat (1992), "The use of information in decision making: An experimental investigation of the impact of computer-based decision aids," *MIS Quarterly*, 16 (3), 373-393.

Toulas, Bill (2021), "BlackBerry is reportedly selling patents to Huawei," Technadu.com (accessed January 15 2021) [Available at: https://www.technadu.com/blackberry-reportedly-selling-patents-huawei/240632/].

Vessey, Iris (1991), "Cognitive fit: A theory-based analysis of the graphs versus tables literature," *Decision Sciences*, 22, 219-240.

Vimalnath, Pratheeba, Anjula Gurtoo, and Mary Mathew (2018), "Patent characteristics and the age-value relationship: Study of Ocean Tomo auctioned US singleton patents for the period 2006–2008," *R&D Management*, 48 (3), 271-289.

Wu, Qingsheng, Xueming Luo, Rebecca J. Slotegraaf, and Jaakko Aspara (2015), "Sleeping with competitors: The impact of NPD phases on stock market reactions to horizontal collaboration," *Journal of the Academy of Marketing Science*. 43, 490-511.

Yang, Fan, Yi Qian, and Hui Xie (2024), "Addressing endogeneity using a two-stage copula generated regressor approach." *Journal of Marketing Research*, 62 (4), 601-623.

Web Appendix

W.1 Data Descriptions

Figure W1: Empirical Density Functions of Log(Estimate) for Sold and Unsold Patents Pre- and Post-Treatment

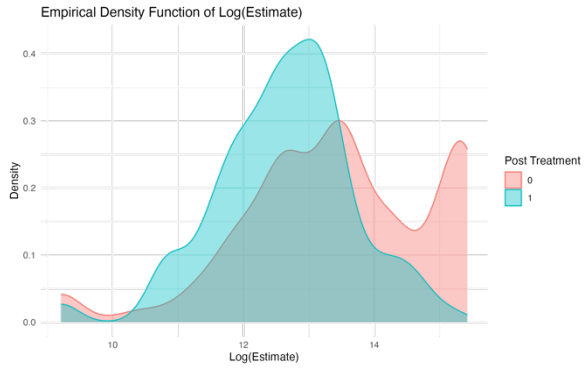


Figure W2: Histograms of the Number of Potential Licensees for Sold and Unsold Patents Pre- and Post-Treatment (Excluding Patents in Auctions 1 and 2)

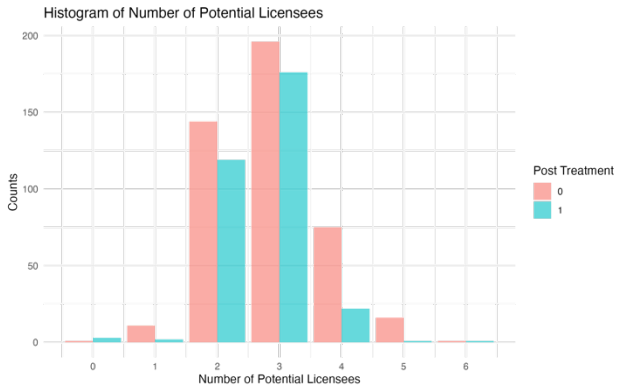
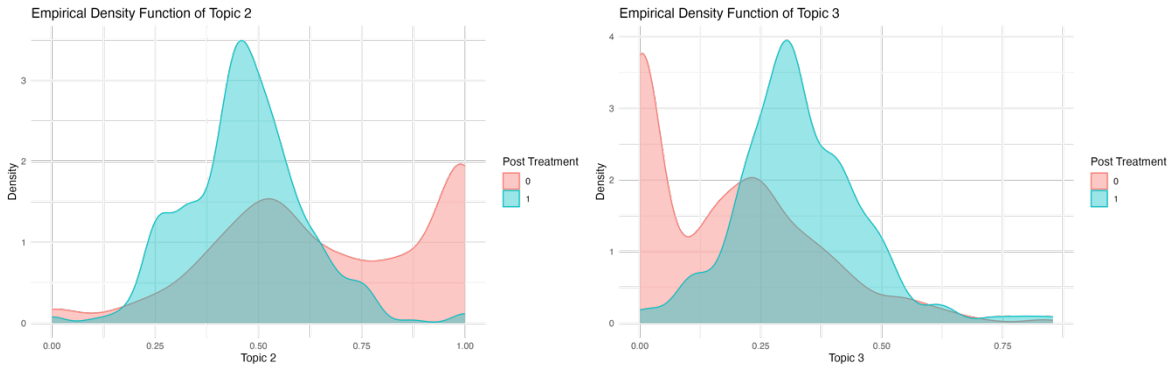


Figure W3. Empirical Density Functions of Weights on Topic 2 (Left) and Weights on Topic 3 (Right) for Sold and Unsold Patents Pre- and Post-Treatment



W.2 Evidence of Improvement in Information Content

In the main paper, we have demonstrated tangible improvements in the presentation of information, making it more salient to audiences, particularly potential buyers. In the following analysis, we further provide evidence that the treatment enhanced the information content conveyed in the text, based on estimates from the Heterogeneous Supervised Topic Model (HSTM) in the web appendix. Since each text description in the catalog corresponds to a specific auction lot, we conduct the following analysis at the auction-lot level rather than the patent level.

For each auction-lot description, we compute the narrative score by summing the estimated coefficients of individual words (i.e., tokens) from the HSTM model, weighted by the document-word frequency vector. We further scale the narrative score such that it ranges between 0 and 1. We also classify auction lot text descriptions into two distinct groups based on the narrative score using a predefined threshold. If the narrative score exceeds the threshold, the text is categorized as having an overall pro-sale tone; otherwise, it is assigned to the passive tone group. The threshold is set at the 40th percentile of the narrative scores across all patent lot descriptions, aligning with the empirical distribution of sales, as approximately 40% of the 509 auction lots in our sample were successfully sold.

Table W1 reports the regression results for the narrative score and the likelihood of a text file exhibiting a pro-sale tone based on the treatment and auction lot characteristics (averaged across patents within each lot). For the binary outcome variable indicating a pro-sale tone, we employ a probit model. Additionally, we control for technology category fixed effects and season fixed effects to account for potential variations across different technology classes and auction seasons.

As can be seen, the coefficient of the post-treatment period dummy is positive and statistically significant for both outcome variables at the 1% level. Holding all other characteristics

constant, the post-treatment narrative score exhibits an 18.5% increase, indicating a substantial improvement in the information content conveyed in the auction lot descriptions. We also find that a higher narrative score is associated with a greater number of forward citations but a smaller number of patents per lot. This suggests that patents with more forward citations tend to have higher-quality descriptions, potentially reflecting their greater technological significance. Conversely, larger patent bundles may make it more challenging to craft effective and engaging descriptions, as increasing the number of patents in a lot could dilute the clarity and focus of the narrative.

Table W1: Effects of Treatment and Auction-Lot Characteristics on Narrative Score and Overall Pro-Sale Tone in Auction Catalog Descriptions

	Narrative score		Overall pro-sale tone (binary)	
Post	0.185***	(0.049)	0.485***	(0.140)
I. Featured characteristics				
Log(Estimate)	0.025	(0.017)	0.088*	(0.052)
No. potential licensees	0.025	(0.017)	0.068	(0.049)
II. Citation metrics				
Forward citations	-0.002	(0.002)	-0.003	(0.006)
Weighted forward citations	0.001**	(0.001)	0.004**	(0.002)
Backward citations	0.002	(0.001)	0.004	(0.004)
III. Other Characteristics				
No. patents in a lot	-0.011**	(0.005)	-0.066**	(0.026)
Branded	0.062	(0.045)	0.219*	(0.132)
Patent age	-0.002	(0.006)	-0.004	(0.018)
VII. Fixed effects				
NAICS category	Y		Y	
Season dummy	Y		Y	
Number of auction lots	506		506	

Notes: The narrative score and overall pro-sale tone of a text description are measured using the estimated word coefficients from the HSTM model, which was trained on 20 topics. These estimates are based on the seed that achieved the highest ROC AUC score of 0.68, where the top 65 overall pro-words and passive words identified by the HSTM model are presented in Table 9 of the main paper. For each auction lot, the characteristics are computed by average over the patents pooled in the lot. The numbers in brackets are estimated standard errors.

W.3 Robustness Check: Lead Time Variation

In Table 6 of the main paper, we document substantial effects of information presentation on patent sales. A potential concern is whether these effects are sensitive to the time available for buyers to gather information prior to the auction. In practice, however, most auctions in other product categories provide advance notice and lead times comparable to those in Ocean Tomo patent auctions, with vehicle auctions being an exception.

To address this concern more directly, we conduct a robustness analysis excluding IPs from the 2006 Fall Second Auction. Our data include only nine text descriptions for lots in this auction, which is considerably fewer than in the other auctions, where each contains more than 50 lot descriptions. This auction was held unexpectedly, in addition to the regularly scheduled Spring and Fall auctions, and may therefore have allowed buyers less preparation time. If limited lead time materially attenuated the role of information presentation, the estimated effects should *weaken* in this restricted sample.

Table W2 reports the 2sCOPE estimated effects of information presentation on patent sales or not, both without and with heterogeneous treatment effects. The results are quite similar to those in Table 6. This evinced that variation in lead time or advance notice does not drive our findings, and that our estimated effects of information presentation are robust to differences in buyers' preparation windows.

Table W2: 2sCOPE Estimated Effects of Information Presentation on Patent Sales (Excluding IPs in the 2006 Fall Second Auction)

	(1)		(2)	
	Estimate	SE	Estimate	SE
Post	0.443	(0.222)	0.116	(0.289)
1. Seller's estimated IP value				
Log(Estimate)	-0.081	(0.043)	-0.074	(0.032)
2. Features				
No. potential licensees	0.233	(0.134)	0.222	(0.109)
Weight on Topic 2	-0.669	(0.350)	-0.617	(0.288)
Weight on Topic 3	-0.141	(0.463)	-0.493	(0.312)
3. Citation metrics				
Forward citations	0.003	(0.002)	0.002	(0.001)
Backward citations	0.0003	(0.0007)	0.001	(0.0006)
4. Other Characteristics				
No. patents in a lot	-0.020	(0.011)	-0.018	(0.008)
Branded	0.024	(0.031)	0.030	(0.021)
Patent age	-0.001	(0.004)	0.007	(0.004)
5. Textual characteristics				
Log(Word count)	0.141	(0.094)	0.079	(0.050)
Semantic similarity	-0.053	(0.038)	-0.030	(0.022)
6. Heterogeneous treatment effects				
Post × Log(Estimate)			0.027	(0.023)
Post × Forward citations			0.001	(0.001)
Post × Backward citations			-0.005	(0.003)
Post × No. patents in a lot			0.031	(0.015)
Post × Patent age			-0.026	(0.012)
7. Auction-specific effects				
Max Log(Estimate)	0.379	(0.191)	0.306	(0.143)
Avg. No. potential licensee	-0.138	(0.100)	-0.091	(0.066)
Avg. forward citations	-0.024	(0.013)	-0.025	(0.011)
Avg. backward citations	0.038	(0.022)	0.038	(0.020)
Avg. patent age	-0.082	(0.046)	-0.066	(0.033)
Correlations with estimated structural errors				
Log(Estimate)	0.131	(0.066)	0.140	(0.062)
No. potential licensees	-0.319	(0.173)	-0.311	(0.149)
Weight on Topic 2	0.088	(0.049)	0.079	(0.039)
Weight on Topic 3	0.076	(0.108)	0.152	(0.079)
No. observations		1,330		1,330

Notes: The effects of information presentation on patent sales are estimated based on 1,330 patents in seven auctions with patent NAICS category-fixed effects, excluding IPs in 2006 Fall Second Auction. "Post" is a dummy variable, equaling one for patents offered in auctions after 2007. For other explanatory variables, see Tables 2, 3, and 4. Numbers in brackets are estimated standard errors, where bootstrap is used for all 2sCOPE estimates.

W.4 Extending Linkage Principle to Include Information Presentation

Let $p(\overline{v}, v)$ be the expected payment by bidders with value v, and with bid value $\overline{v}$. Now, let A and B are two auctions of the same item, but with two different information presentation formats. Information in auction A are organized, and information in auction B are disorganized. Furthermore, the winner pays $P=P_H$ when perceived risk of items is high, and pays $P=P_L$ when the perceived risk of items is low with $P_L \geq P_H$. Now assume that, in each auction, there is a symmetric and strictly increasing equilibrium bidding function with $p^A(\underline{v}, \underline{v}) = p^B(\underline{v}, \underline{v}) = 0$. Given information is organized, i.e., $\lim_{X \to X_O} p(\overline{v}, v) = P_L$, $p^A(\overline{v}_a, v_a) \geq p^B(\overline{v}_b, v_b)$. The seller's expected revenue in auction A is at least as large as the expected revenue in auction B.

Alternatively, we extend Theorem 8, (Milgrom and Weber 1982: 1102) to include information presentation effects on seller's revenue. In this case, let R_{DO} be the expected selling price when the information X_{DO} is disorganized, and let R_O be the expected price when information X_O is organized. Consider the following expected selling prices:

$$R_{DO} = E\left[v_{do}(Y_1, Y_1; X_{DO}) \mid \{X_1 > Y_1\}\right],$$
$$R_O = E\left[v_o(Y_1, Y_1; X_O) \mid \{X_1 > Y_1\}\right]$$

Given that revealing structured information reduces risk, $R_O \geq R_{DO}$. Please see Theorem 8 for the above proof.

W.5 Validation: Laboratory Experiment

To understand the underlying decision processes at different information presentation and validate our findings from the natural experiment, we performed a laboratory experimental study. We hypothesize that information presentation format (salience and organization) will affect perceived complexity, attitude toward the product, perceived quality, perceived risk, perceived cognitive load, and perceived information sufficiency for a purchase decision, which in turn will affect the final price of the item. We first identify three types of products (search, experience, and credence), then conduct a pre-test to identify which information is salient for making a purchase decision for each product type and confirm information organization format. We performed a between-subjects study manipulating information salience (yes, no) and organization (organized, disorganized) across the product descriptions. Information salience is operationalized using the pre-test results. Information organization is operationalized by presenting information in essay format or bullet-point format, and the manipulation was pre-tested. Stimuli, including information presentation sequence, information content, font type and color, and image size are held constant across the product descriptions.

For the experiment, responses were solicited for voluntary participation from undergraduate students registered as business majors at a major south-western university in the United States. Participants earned credit for taking part in the study. In all, 167 participants completed the study. Each participant viewed four products. All participants were under 50 years of age (19 – 48). The mean age was 21.5 years. Males consumed 64.1% of the sample, and 31.1% had a family income of over a $100,000 per year.

Stimuli. Participants rated their perceptions of product information presentation using tasks specifically designed for this study. The products and descriptions for the stimuli were

sourced from various internet websites. All product and brand names were removed from the pictorial and textual stimuli. To determine if effects of information presentation on value perception are specific to certain product categories, we selected four types of products, specifically a LED television for an expensive search product, a tablet computer for an inexpensive search product, a Caribbean cruise for an experiential product, and a Professional Certification Program for a credence product. Participants viewed pictures of the products and associated information descriptions in an on-line survey administered through the university's student research program. A sample treatment is shown in Web Appendix Figure W4.

Dependent Variables. The dependent variables complexity (Holbrook 1981), attitude toward the product (Shimp and Bearden, 1982), perceived product quality (Petroshius and Monroe, 1987), perceived cognitive load (Moray, 1982), perceived risk (Shimp and Bearden, 1982), and perceived information sufficiency for purchase (Ratchford, 1987) were measured with items adapted from validated scales. Complexity, perceived risk, and perceived information sufficiency were single item scales. The Cronbach's alpha for each of the multi-item scales (attitude toward the product, perceived product quality, perceived cognitive load) were 0.80, 0.84 and 0.80 respectively.

Independent Variables. All subjects completed surveys utilizing the four product types. The independent variables organization (organized/disorganized), and salience (important/unimportant) were manipulated for each product type. Information organization was manipulated by presenting the textual information in bullet point or paragraph format. Information salience was manipulated by presenting unimportant and important information.

Results. An ANOVA of salience and organization of information presentation revealed that information presentation showed a significant effect on all factors except perceived risk.

Figure W4. Information Presentation Stimulus Disorganized (left) Organized (Right) with Same Length

Are you ready to leave winter behind and to join us for a Caribbean cruise to the Bahamas?

Let us tell you about a few of the Bahama highlights: The living is easy on The Bahamas islands, a glittering necklace of sand dotting the blue Atlantic. A Caribbean Cruise to The Bahamas drops you into a sun-splashed seascape of 700 islands and 2,000 cays offering idyllic tropical getaways for sailing, snorkeling, fishing, strolling pearly pink-sand beaches, and swimming crystalline turquoise seas.

On Caribbean Cruise Lines, our ships are built with one goal in mind: to make sure that every time you walk up the gangway, you get the sense that you're crossing over into a whole new world of fun.

Once on board, let the good times roll because just about everything is included. Enjoy delicious meals in various onboard restaurants, including your comfy accommodations with the complimentary 24-hour stateroom service. Pull an all-nighter — or as much of one as you can manage — as you revel in the awesome nightly entertainment; then wonder at all the fun there is to have the next day.

10 Day Cruises from Miami to Freeport are $2,000 for a Stateroom with all meals included. December 2011 availability is limited so book today!



Are you ready to leave winter behind and to join us for a Caribbean Cruise to The Bahamas?

The living is easy on The Bahamas islands:

- a glittering necklace of sand dotting the blue Atlantic
- a sun-splashed seascape of 700 islands and 2,000 cays
- an idyllic tropical getaway for sailing, snorkeling, and fishing
- stroll pearly pink-sand beaches, and swim crystalline turquoise seas.

.

On Caribbean Cruise Lines, our ships are built with one goal in mind:

- to make sure that every time you walk up the gangway, you get the sense that you're crossing over into a whole new world of fun.

Once on board, let the good times roll because just about everything is included:

- enjoy delicious meals in various onboard restaurants
- including your comfy accommodations with the complimentary 24-hour stateroom service.

Pull an all-nighter — or as much of one as you can manage:

- revel in the awesome nightly entertainment
- then wonder at all the fun there is to have the next day.

10 Day Cruises from Miami to Freeport are $2,000 for a Stateroom with all meals included.

December 2011 availability is limited so book today!



Information salience was significant for: attitude toward the product: $F(1, 665) = 7.54$, $p < 0.05$;perceived product quality: $F(1, 665) = 9.21$, $p < 0.05$ and perceived information sufficiency

for purchase: $F(1, 665) = 7.63$, $p < 0.05$. Information organization was significant for product complexity: $F(1, 665) = 5.76$, $p < 0.05$; attitude toward the product: $F(1, 665) = 4.66$, $p < 0.05$; and perceived cognitive load: $F(1, 665) = 5.16$, $p < 0.05$. Although we did not find effects on perceived risk and differences across the product types, we did find moderating role of risk in the relationship between organization and attitude towards the product and perceived product quality. Detailed results can be obtained from the authors upon request.

Web Appendix References

Holbook, Morris B. (1981), "Integrating compositional and decompositional analyses to represent the intervening role of perceptions in evaluative judgments," *Journal of Marketing Research*, 18, 13-28.

Petroshius, Susan M. and Kent B. Monroe (1987), "Effect of product-line pricing characteristics on product evaluations," *Journal of Consumer Research*, 13, 511-519.

Ratchford, Brian T. (1987), "New insights about the FCB grid," *Journal of Advertising Research*, 27, 24-38.

Shimp, Terence A. and William O. Bearden (1982), "Warranty and other extrinsic cue effects on consumers' risk perceptions," *Journal of Consumer Research*, 9, 38-46.