Infringing Use as a Path to Legal Consumption: Evidence from a Field Experiment*

Hong Luo[†] Julie Holland Mortimer[‡]

June 25, 2020

Abstract

Digitization has transformed how users find and use copyrighted goods, but many existing legal options remain difficult to access, possibly leading to infringement. In a field experiment, we contact firms that are caught infringing on expensive digital images. Emails to all firms include a link to the licensing page of the infringed image; for treated firms, we add links to a significantly cheaper licensing option. Making infringers aware of the cheaper option leads to a fourteen-fold increase in the ex-post licensing rate, albeit from an extremely low baseline for the control firms. Two additional experimental interventions, designed to reduce search costs for (i) price and (ii) product information, also have large positive effects. Our results suggest that ex-post monetization (e.g., licensing after use) may expand the market, and that rights holders can create value by minimizing search and transactions costs.

*The authors thank Tobias Kretschmer, Michael Luca, Shane Greenstein, Felix Oberholzer-Gee, Martin Watzinger, and seminar participants at Harvard Business School, Temple University, and Advances with Field Experiments 2019 conference for their valuable feedback. We thank the Agency for providing the proprietary data and for its collaboration on the experiments. We also thank Esther Yan and Louise Yu for their excellent research assistance. All errors and omissions are ours. AEA RCT Registry-0003238.

[†]Harvard Business School; email: hluo@hbs.edu.

[‡]Boston College and NBER; julie.mortimer.2@bc.edu.

1 Introduction

Digitization has profoundly transformed the production, distribution, and diffusion of copyrighted goods (Varian (2005); Smith and Telang (2016); Waldfogel (2018)). Use has expanded, but, at the same time, competition has intensified, and sources of potential revenue for creators have become more fragmented. As a participant in the music industry observed, "These days it's essential to collect every income stream one can."[1] Indeed, even established creators may not be able to rely on a small number of high-value transactions; small-value and widely dispersed usage, collectively, can comprise a substantial portion of revenue. As a result, a key challenge that rights holders face in many creative industries is to develop 'digital-age' technological and logistical infrastructures, as well as suitable licensing and monetization models that can track a large and dispersed quantity of usage and enable efficient flows of small streams of money.

For users, general-use search engines and social media have made it easy to search for digital content or to encounter it serendipitously. In contrast, many existing legal options for licensing content remain difficult to access. For example, while images are easy to find on the Internet, it is usually not clear who owns the copyright on an image or where and how to license it legally. For music, a large share of end consumption is well managed by platforms such as iTunes and Spotify, but rights for commercial use and follow-on production are notoriously fragmented.[2] Search and transactions costs matter critically for digital content because the potential value of any individual transaction is relatively small. Thus, even small frictions or inconveniences may be sufficient to deter licensing and, potentially, lead to copyright infringement.[3]

These technological and market dynamics suggest the possibility of an 'ex-post' approach to monetization for copyrighted goods that may complement, not supplant, the ex-ante licensing market. For example, rights holders may approach infringing users to (i) clarify their legal obligations; (ii) increase their awareness of alternative purchasing options (if, for example, the infringed products are too expensive); and (iii) mitigate their search and transactions costs. An ex-post market may create value because it relaxes the requirement to obtain permission prior to use and allows consumers to find products in the manner that is the most convenient for them. Moreover, infringement may provide valuable information, as consumers' use of content is informative about their level of demand and unobserved preferences for specific products. Given that small-volume users tend to be numerous and dispersed, and that predicting demand for creative content is difficult, monitoring infringing use may be a cost-effective means to identify potential customers.

[1] Mark Beaven, in "Going to the Ends of the Earth to Get the Most Out of Music," June 8, 2015, *New York Times*.

[2] Recently, several publishers sued the fitness technology company Peloton for $150 million for failing to license their music (https://techcrunch.com/2019/03/20/peloton-hit-with-150-million-music-licensing-suit/). Even for end consumption of music, certain product types, such as the classical genre, are not suited to the streaming infrastructures that are built for pop songs (https://www.nytimes.com/2019/06/23/business/media/stream-classical-music-spotify.html).

[3] Often, users do not have a clear idea about their legal obligations for licensing or perceive the legal threat as small (Luo and Mortimer (2017)).

Naturally, from a copyright holder's perspective, the value of ex-post monetization is likely to depend on a number of factors, such as the extent of frictions preventing ex-ante legal consumption; rights holders' abilities to capture value ex-post (through either direct licensing or indirect methods such as ad placement); and the cost of monitoring infringing use.

Our paper explores the feasibility of an ex-post monetization approach in a specific setting: the stock-photography (pre-shot images) industry.[4] On behalf of photographers, stock-photo agencies license images to business customers. The industry is divided into premium and micro-stock segments, and there are two types of licensing models: (i) a 'rights managed' (RM) license restricts the use of an image within a pre-specified scope, including duration, purpose, and placement; and (ii) a 'royalty free' (RF) license allows the licensee to use the image without restrictions for a one-time payment. The highest-quality images in the premium segment are licensed through a RM model. This type of product targets usage occasions such as large advertising campaigns, and the price of a license is usually in the hundreds or even thousands of dollars. Obtaining a price quote requires a user to fill out a form specifying the scope of use, and obtaining the price for a different use requires filling out a new form. Remaining premium images are licensed through a RF model at prices of tens to hundreds of dollars for unrestricted use. All images in the micro-stock segment are licensed through a RF model. For small firms that want to display images on their websites, for example, the micro-stock segment tends to be the preferred option because of both its low price (tens of dollars or less) and lack of restrictions on use. Over the past decade or so, the stock-photography industry has experienced the trends described above: (i) increasing expansion of image use (both legal and illegal); and (ii) a sharp decline in the share of revenue that comes from high-value transactions.[5]

The data used in this paper are generated from a new field experiment conducted by one of the leading stock-image agencies (hereafter, the 'Agency') that offers products in both the premium and micro-stock segments through separate websites. The Agency monitored unauthorized use of its RM premium images—the most expensive images, representing a small percentage of its total portfolio—by commercial websites. The experiment focuses on infringement cases involving the smallest firms, for which the infringed RM premium (hereafter, premium) image is likely to be too expensive.[6]

The Agency's goal was to encourage infringing users to purchase a legitimate image license by, first,

[4]Glückler and Panitz (2013) estimates that the global revenue of the stock-photography industry was \$2.88B in 2011. In comparison, the revenues of ASCAP and BMI, the two largest performance rights organizations that collect royalties on behalf of copyright holders in the music industry for public use of their works, were \$1B each in 2012.

[5]For a leading agency in this industry, for example, RM premium images—which represent a very small proportion of its portfolio in terms of quantity—used to account for about 40 percent of its total revenues 15 years ago. As of 2018, the revenue share of this type of product dropped to ten percent. Multiple factors may account for these changes, including increasing competition from low-cost providers that focus on the micro-stock segment and changing advertiser behaviors (for example, moving away from traditional channels such as billboard displays towards digital advertising).

[6]These cases make up about 20 percent of all detected cases infringing on the Agency's RM premium images.

increasing their awareness of the significantly cheaper micro-stock website that is also owned by the Agency and, second, lowering search costs for relevant product and price information. Specifically, the experiment includes two control and four treatment groups. Emails to all the treatment groups offered a micro-stock licensing option in addition to the option of licensing the infringed premium image. In contrast, emails to the control groups included only the premium licensing option. We further varied the treatment conditions in two dimensions. The first intervention intends to reduce the search costs for product information; in particular, emails to two treatment groups recommend four images from the micro-stock site that are similar to the infringed image, whereas emails to the other two treatment groups contain only a link to the home page of the micro-stock site. The second intervention intends to reduce the search costs for price information—emails to two treatment groups and one control group add the price information of the infringed premium image, thus making the price comparison between the two options immediately clear. (The micro-stock price is available in the emails to all treatment groups.)

We find that awareness of the micro-stock option leads to a fourteen-fold increase in the probability of licensing, albeit from an extremely low baseline for the control groups. The treatment groups' average licensing rate, 2.63 percent of opened emails, consists almost entirely of licenses of the micro-stock option, and is several times higher than the most effective email marketing campaigns conducted by the Agency for the same type of products. This result is consistent with the idea that infringing use is informative of demand—at the basic level, the fact that these users are using images on their websites suggests that they may, on average, have a higher willingness to pay than a generic or even somewhat targeted group of non-customers.

In the treatment groups, both additional interventions have a large, positive average effect: image recommendation increases the probability of licensing a micro-stock image by 40 percent, and the price comparison information increases the micro-stock licensing rate by 27 percent. Results on the effects of the interventions on the likelihood of searching the two websites suggest that both interventions are effective because they ultimately reduce the costs of finding a replacement image. However, the precise mechanisms and the types of users affected by the two interventions appear to be different: by reducing the cost of finding a replacement image directly, image recommendation mainly induces marginal users (i.e., those with a relatively low willingness to pay) to start a search instead of choosing the outside option; however, providing premium-price information may direct infra-marginal users (i.e., those with a relatively high willingness to pay) away from the premium option faster, thereby leaving them more time to search the cheaper site.

Our findings suggest that ex-post monetization may expand the market to include consumers who would not have been active in an ex-ante market, and that rights holders may create value in the ex-post market by minimizing search and transactions costs. Though the specific approaches that rights holders can take may differ by context, our results provide evidence for the following features that may characterize copyright markets more generally: (i) users face search and information frictions in the ex-ante market; (ii) infringing use can reveal information about consumer demand and preferences for specific products; and (iii) simplifying the search and transaction process is important for small-value transactions. Admittedly, these results, alone, do not address the likely profitability of an ex-post monetization approach generally or how it compares with alternative approaches (e.g., reducing the search and transactional frictions in the ex-ante market), both of which are probably context-dependent. Furthermore, the results do not permit us to analyze potential equilibrium responses by users and rights holders, or changes in the nature of competition.

1.1 Related literature

We are aware of two theoretical studies on the impacts of allowing users to use copyrighted products first and compensate rights holders afterwards. Using an incomplete-contract framework, Gans (2015) argues that an alternative copyright regime that allows follow-on products to be created without obtaining permission from right holders—who are compensated later—can mitigate transactions costs and incentivize both follow-on and original investments.[7] In a model without any search or transactions frictions, Hua and Spier (2019) show that firms can generate higher profits by committing to a 'soft' out-of-court settlement policy that encourages high-type consumers to purchase the product and low-type consumers to pirate. Upon being caught, infringers can settle and license the product for future use, thus facilitating price discrimination and expanding the market.[8] Our paper complements these theoretical studies using a different approach: we conduct a field experiment to test the feasibility of ex-post monetization. The experiment focuses on search frictions—it provides evidence for their presence in the ex-ante market and shows that rights holders can create value by mitigating search frictions in the ex-post market.[9]

[7]Gans (2015) focuses on transactions costs due to hold-up problems, because the amount of remix that rights holders (i.e., creators of the original content) will allow to be published and the compensation to rights holders can only be negotiated after investments are made by the follow-on and original creators. The paper examines different compensation policies to original creators (fixed payment or revenue-sharing) and compares them to both the traditional copyright regime that allocates the control rights to original creators and a no-IP regime that allocates the control rights to follow-on creators. Menell (2014, 2016) make a legal case for a remix compulsory licensing regime that aims to reduce various types of transactions costs in the digital age.

[8]See Peitz and Waelbroeck (2006a) and Belleflamme and Peitz (2012, 2014) for comprehensive reviews of the theoretical literature on piracy. This literature also points to a number of reasons that piracy or weak IPR enforcement may increase firm profits, via, for example, a stronger network effect that enables rights holders to charge a higher price for users who do not infringe (Conner and Rumelt (1991); Takeyama (1994); Shy and Thisse (1999)) and a bundling effect of product sharing that reduces consumer heterogeneity (Bessen and Kirby (1989); Bakos et al. (1999)). These theories are mostly about indirect monetization through charging more to the legal users in the ex-ante market and do not explicitly consider the monetization of the specific infringement incidences. Another positive effect of piracy is the sampling effect (Peitz and Waelbroeck (2006b)) that reduces product uncertainty for infringers (who may also be future customers), which we discuss in the next paragraph.

[9]For patents, settlement agreements are often combined with licensing agreements that allow the infringing firms to license the product in the future. The literature on patent litigation typically focuses on disputes between competing firms, but recent studies on non-practicing entities have been about enforcement strategies against users of patented technologies or products embodying such technologies (Choi and Gerlach (2018) and Cohen et al., (forthcoming)). It is important to note that the approach we propose is qualitatively different from strategies that aim to maximize enforcement revenues by exploiting legal threats and ambiguous infringement situations. Patent rights often have uncertain boundaries or questionable validity (Bessen and Meurer (2006)), and

Our paper also contributes to two strands of the empirical literature on piracy. First, it joins a small number of recent studies that examine the effects of improved availability and affordability of legal options (Danaher et al. (2010); Aguiar and Waldfogel (2018)).[10] Using a field experiment, Godinho de Matos et al. (2018) randomly gift some infringing households with free access to ten channels streaming movies and TV shows. They find that, on average, the gift increases overall TV consumption but has no effect on the households' likelihood of accessing pirated content via BitTorrent. Our paper adds new field-experimental evidence on proactive supplier efforts to convert infringing users to legal paying customers; it estimates the causal effects of reducing two types of search costs without the use of price subsidies.[11] Furthermore, while the literature focuses on consumer piracy, our results shed light on behaviors of firms who infringe for commercial purposes. Second, our paper relates to the 'sampling effect' as it is understood in the literature—piracy may encourage users to experiment and discover products or artists that are new to them, which can subsequently lead to the purchase of similar products or other products by the same artists (e.g., Peitz and Waelbroeck (2006b); Oberholzer-Gee and Strumpf (2007); Smith and Telang (2009); Gans (2012); Zhang (2016); Kretschmer and Peukert (2017)).[12] While the literature on the sampling effect focuses on consumers' learning about product quality, our experiment is about producers' learning about user preferences revealed by infringement and their use of this information to exogenously reduce user-specific search costs for replacement products.

Our results also relate to the literature on information provision and search costs, in particular in the context of electronic marketplaces (Alba et al. (1997); Bakos (1997)). Prior work focusing on competition has shown that lowering search costs for product characteristics may have different impacts on competition than lowering search costs for price information (e.g., Brynjolfsson and Smith (2000); Lynch and Ariely (2000)). Our paper shows that a firm's ability to simplify search for both price and product information across its full range of products may increase a user's expected payoff from consuming the firm's products relative to the outside option. Ultimately, both treatment interventions in our experiment appear to have facilitated the matching of users to their preferred products. From this perspective, our paper also relates to studies on the effect of information provision on the quality of matching (Anand and Shachar (2011);

it may be hard for patent holders to commit ex-ante to a licensing price that is deemed fair and reasonable (Shapiro (2010); Scott Morton and Shapiro (2014)).

[10] Exploiting natural experiments, Danaher et al. (2010) and Aguiar and Waldfogel (2018) show that the availability of content through digital channels such as iTunes and Spotify significantly influences piracy.

[11] The proactive use of infringement information by copyright holders to identify potential customers and to serve them better seems a qualitative change of mindset for practitioners in creative industries. Not long ago, infringement information was used primarily as a way to identify targets for lawsuits; to send take-down notices; to pursue settlement for past infringement without providing a path towards licensing for future use (Bhattacharjee et al. (2006); Danaher et al. (2014); Reimers (2016); Luo and Mortimer (2017)); or, more typically, it was not utilized at all.

[12] In the debate over whether piracy displaces or complements sales, prior studies generally find that piracy hurts sales on net (e.g., Rob and Waldfogel (2006); Danaher and Smith (2014); and see Waldfogel (2012) for a comprehensive survey).

Tadelis and Zettelmeyer (2015)) and the effects of reducing search costs on peer-to-peer platforms. (See Fradkin (2017) and Horton (2019) on providing better-updated information about the availability of homes on Airbnb or of workers on oDesk.)

Finally, our paper relates to the literature on recommender systems that help users to discover new products and to deal with information overload (Resnick and Varian (1997)). While this literature (especially studies in computer science and information systems) focuses mainly on system designs, important economic considerations include the effects on the diversity of sales (Fleder and Hosanagar (2009)); potential crowding-out effects on non-recommended products or workers (Horton (2017)); and incentive mechanisms for the provision of evaluations (Avery et al. (1999)).[13] Recommender systems have been applied to a wide range of settings, especially in the e-commerce, entertainment, and news industries.[14] Our paper provides a novel application of copyright management in which the information on user preference is revealed from illicit behaviors. Our paper also contributes new field-experimental evidence on the causal effect of a recommender system (Jannach and Hegelich (2009); Belluf et al. (2012); Sharma et al. (2015); Carmi et al. (2017)), which is often challenging to identify because recommendations and consumer interest are expected to be correlated. Our results show that only ten percent of all recommended purchases would have happened without the recommender system. This result provides an interesting contrast to Sharma et al. (2015) who show that in the context of Amazon.com, at least 75 percent of recommendation click-throughs would have happened even without Amazon's recommender system.

2 Experimental Design

Table 1 summarizes our experimental design: there are two control and four treatment groups. Emails to the two control groups (the first row in table 1) include only a link that directs the user to the licensing page of the infringed premium image currently displayed on the firm's website. Emails to all the treatment groups add an affordable micro-stock licensing option (indicated by "Micro" in the second and the last rows), in addition to the premium option.

We further vary the treatment conditions in two dimensions, aiming to reduce search costs for relevant product and price information. First, emails to two treatment groups (second row) contain a link to the home page of the micro-stock site. In contrast, emails to the other two treatment groups (the last row, indicated by "Rec") recommend, based on the Agency's proprietary algorithm, four images from the micro-stock site that are similar to the infringed premium image. The emails contain a thumbnail and a link to the licensing page

[13] See Adomavicius and Tuzhilin (2005) for a survey of the design of recommender systems, which are categorized into content-based systems that recommend items similar to those that a user liked in the past; collaborative filter-based systems that recommend what similar customers bought or liked; and hybrid methods.

[14] See Lu et al. (2015) for a comprehensive review of the applications of recommender systems.

for each of the four recommended images. This treatment is likely to be helpful in finding a replacement image, as there are millions of images on the website. It may not be obvious what keywords can be used to narrow down the search, and the potential multitude of choices may be overwhelming.

Second, emails to two treatment groups and one control group (indicated by "Price" in the second column) add the following premium-price information: "Licensing costs for online use of Rights Managed images typically range from $545 to $1140 for a 3-month period." The price information on a micro-stock image ("as low as $12 per image") is presented in the same way in the emails to all four treatment groups. As mentioned earlier, obtaining the price of a RM premium image for each different use specification requires filling out a new form; and to have a sense about the typical price range for the right specifications may take a few tries. Providing such information in the emails is designed to make the price comparison of the two product types immediately clear.

For illustration purposes, figure 1 presents the email template used for one of the treatment groups ("Micro + Rec"). The itemized portion in the middle is the only part that varies across groups.[15] Emails to the corresponding control group ("0") in the same column of table 1 do not contain the third bullet point; that is, they include only a link that directs the user to the licensing page of the infringed premium image. Also in the same column of table 1, emails to the other treatment group that do not recommend images ("Micro") include the third bullet point but stop at "$12 per image." Finally, the templates used for groups given premium-image price information (i.e., the three groups in the second column of table 1) are exactly the same as their counterparts in the previous column, except that the premium-price information described in the previous paragraph is added to the end of the second bullet point.

Cases included in the experiment involve small businesses whose infringement (of premium images) was identified within the preceding two years. Cases were allocated across groups using a random-number generator. We intentionally allocated more cases to the two treatment groups for which similar images were recommended, as, ex-ante, the Agency deemed this to be the most constructive approach. We sent 24,090 emails in four batches between November 15 and December 04, 2017.[16]

2.1 Analysis sample

To generate the analysis sample, we removed cases for which (1) the email was bounced back due to an invalid address (13 percent); and (2) the email was not opened in the first 14 days (65 percent of the valid emails). Appendix table A3 confirms that both likelihoods are statistically the same across groups. We focus on outcomes in the first 14 days to avoid the confounding effect that worse-performing groups in the initial

[15] Appendix figures B1-B5 present the email templates for the other five groups.

[16] Appendix B provides more details about the experiment. Appendix table A1 summarizes the allocation of cases across groups and across batches, and table A2 shows that the groups are well-balanced with respect to basic case characteristics that we explain in the next section.

two weeks would receive more intense follow-up, as only cases that had not yet licensed received follow-up emails after 14 days. Our data show that about 65 percent of licensing takes place within the first three days, and 79 percent takes place within the first 14 days. The final analysis sample consists of 7,407 cases.

Table 2 shows that the groups are well-balanced with respect to basic case characteristics. Overall, 12 percent of the cases involve the unauthorized use of multiple images represented by the Agency. For multi-image cases, the email displays only one of the images without mentioning that it is a multi-image case. Across all cases, 68 percent of the displayed images are high resolution, and 80 percent are displayed on a secondary page of the firm's website rather than on the home page. The average age of the cases (the number of months between the date when the case was identified and the date when the email was sent) is 14.5 months. It is possible that the infringing images would no longer be on display at the time of our experiment, as some of the firms that responded to our emails pointed out.[17] The raw data show a small but marginally significant negative correlation between the age of a case and the likelihood of licensing, which is consistent with the idea that the older the case, the more likely that the images were no longer on display.[18] It is, thus, possible that the overall licensing rate would be higher if emails were sent in a more timely manner. The final column of table 2 reports the log of the number of total stock images on a website. This variable captures the number of unique images identified as stock images represented by several stock-photo agencies, including, but not restricted to, the Agency with whom we ran the experiment. These images could have been licensed through legal means or used without authorization. Though noisy, this variable is the only systematically available variable that may be a proxy for a firm's overall demand for images.

3 Results

3.1 Effects of awareness of the micro-stock licensing option

Column 1 of table 3 shows that the average 14-day licensing probability (including both the premium and micro-stock options) for all treatment groups is fourteen times that for the two control groups (2.63 versus 0.17 percent for opened emails, and p-value < 0.001).[19] Almost all of the licenses are of the micro-stock option. Only seven licenses—three in the two control groups and four in the treatment groups—are of the premium option. These results show that the premium option is too expensive for almost all the firms in our sample. Hence, the baseline licensing rate in the control groups is extremely low, and awareness of an

[17] A small number of recipients sent their clarification questions, explanations for the incidences, or further inquiries to the contact email address provided in the emails. These emails received replies within a day or two, following a template that was concise and friendly in tone. Importantly, the reply emails to users were written so as not to interfere with different treatments; in particular, they made no recommendations of the more affordable licensing site or replacement images.

[18] The correlation between the age of a case and the likelihood of licensing is -0.0218, and p-value is 0.0608.

[19] To be conservative, all p-values reported in the paper are based on two-sided tests.

affordable option can significantly increase the licensing rate.

Columns 2 and 3 of table 3 show that because the revenue per premium license is much higher than that for the micro-stock license (\$423 vs. \$23 per license), the expected revenue per case for the treatment groups is 2.5 times that for the control groups (\$0.92 vs. \$0.38, and p-value = 0.218), despite the 14-fold increase in the licensing rate.[20]

Even though the licensing rate by the treatment groups is low, it is important to emphasize that it is several times higher than the rate in the Agency's most effective email marketing campaigns for micro-stock images.[21] This is consistent with the idea that infringing use is informative of demand. The fact that these users are using images on their websites suggests that they may, on average, have a higher willingness to pay than a generic or even somewhat targeted group of non-customers. Furthermore, the ability to observe the specific product chosen by a given user seems important for effective recommendations, given the inherent difficulty in predicting consumer preference ex-ante (especially for creative products). Lastly, it is also important to make the distinction between pre- and post-infringement demand. The higher licensing rate in our context may be partly driven by the fact that these websites are, to some extent, "locked in" to a particular type of images, given other sunk investments made about the website design before the arrival of our emails.

Given that licensing of the premium option is extremely rare, the rest of our analysis focuses on licensing of micro-stock images by the four treatment groups. Specifically, we examine the effects of the two interventions: (i) image recommendations; and (ii) provision of the premium-image price information. We first report the effects on licensing outcomes in section 3.2; then, in section 3.3, we examine the effects on the probabilities that users search the premium and the micro-stock sites, which may shed light on potential mechanisms.

3.2 Effects of the two interventions on licensing the micro-stock option

Table 4 presents OLS regression results of the effects of the two interventions on micro-stock licensing, controlling for case characteristics and batch dummies. Column 1 shows that, controlling for the other intervention, the average effect of image recommendations on the likelihood of licensing is 0.8 percentage

[20]In order to reliably distinguish the effect from zero, we need to increase the sample size substantially. Assuming the same impact-to-standard-deviation ratio, 17,006 observations (2.3 times our current sample size), evenly split between control and treatment conditions is required for a test with a power of 80% at the 5% (one-sided) significance level (Lewis and Rao (2015)). The impact-to-standard-deviation ratio (also known as Cohen's d) is calculated as $\frac{\Delta \bar{y}}{\hat{\sigma}} = \frac{(0.92-0.38)}{14.03} = 0.038$, where 14.03 is the pooled standard deviation from the (pooled) control groups and the (pooled) treatment groups. Of course, because there are only four licenses from the control groups, the impact-to-standard-deviation ratio may not be estimated correctly.

[21]This comparison has incorporated respective email-open rates and, hence, is conditional on all emails sent (not just on opened emails). For confidentiality, detailed statistics about the Agency's marketing campaigns are not disclosed. It is plausible that the relatively high licensing rate in our experiment is driven partly by an implicit threat, even though the emails are generally friendly by design.

points (p-value 0.051). This represents a 40-percent increase, as the licensing probability is 0.020 for the two groups without image recommendations. The average effect of providing premium-price information is 0.6 percentage points (p-value 0.129), or a 27-percent increase relative to the two groups that did not receive price information. Column 2 reports the average effects of the two interventions on the expected revenue per case: the effect of image recommendations is 0.217 (p-value 0.034), representing a 59-percent increase relative to the two groups without image recommendations; and the effect of price information is 0.181 (p-value 0.086), or a 43-percent increase relative to the two groups without price information. For both interventions, the increase in average revenue is driven solely by a higher probability of licensing, rather than by a change in revenue conditional on licensing, as the average revenue per license is about $20 across groups.[22]

Columns 3 and 4 include an interaction term between these two interventions. The interaction term, though not statistically significant, is negative and economically non-trivial for both the licensing likelihood and the expected revenue per case. This suggests that the two interventions are likely to be substitutes for each other; that is, the effect of one intervention is smaller when users are already given the other intervention. As we discuss later, in section 3.3, one potential explanation for this negative interaction effect is that the two interventions appear to influence the licensing outcomes through a similar mechanism—saving users' time and effort to find a replacement image. Thus, when one intervention is in place, the incremental effect of the other is likely to be smaller.

3.3 Search and potential mechanisms

The above results show that both interventions have an economically large effect on licensing outcomes. To better understand potential mechanisms, in this section, we analyze whether a user conducts a search, using data on whether the user clicks the email links to either of the two licensing sites. Before presenting the results, we describe a simple model of user search choices that might help us interpret the data.

Consider a user in group "Micro" who is aware of both licensing options but is not offered either of the two interventions. For the premium option, the user knows which product to use (i.e., the image currently being used) and her willingness to pay for the image, v, but she needs to learn the price. The user has a belief about the distribution of the price, $p \sim F[\underline{P},\overline{P}]$. For the micro-stock option, the user knows the price level ("as low as $12") from the email but needs to search for a replacement image; she is uncertain about the quality of the cheaper images but believes that they will give her a value less than v. Let $v \cdot q$ be the value

[22]The only control variable that is significantly related to licensing probabilities is log(Total stock images on site+1), a noisy proxy for a firm's demand for images; and, intuitively, the coefficient is positive. For each of the two interventions, unreported results (both raw data and regression analysis) show that users with above-median values of log(Total stock images on site+1) drive most of the effects. For example, the raw-data comparisons show that the average effect of image recommendation on the likelihood of licensing is 0.5 percentage points (from 0.016 to 0.021, p-value is 0.403) for users with below-median values of this demand proxy; and it is 1.2 percentage points (from 0.024 to 0.036, p-value is 0.088) for users with above-median values.

of using a micro-stock image; and the user has a belief about the distribution of the quality, $q \sim G[0,1]$.

To obtain either the price or quality information, the user needs to incur a direct search cost; let $c_1 > 0$ be the cost of searching for the premium-price information and $c_2 > 0$ be the cost of searching for a replacement image on the micro-stock site if the user visits the site directly. Let $c_3 \geq 0$ be the extra cost of searching for a replacement image if the user first spends time on the premium site (e.g., due to time constraints). In the following, $c_3 = 0$ serves as a useful benchmark.

We assume that the user has three choices: (1) search for the price information on the premium site first, and, if the price is too high, conduct a second search on the micro-stock site for a replacement image or take the outside option; (2) search the micro-stock site directly; and (3) take the outside option directly.[23] The expected payoffs of the three initial search choices are:

$$U(v) = \begin{cases} U^{\mathrm{P}}(v) = \begin{array}{l} \Pr(v-p \geq m(v))E_p[v-p|v-p \geq m(v)] \\ +(1-\Pr(v-p \geq m(v)))m(v) - c_1 \end{array} & \text{if visiting premium site first} \\ U^{\mathrm{M}}(v) = \Pr(vq \geq 12)E_q[vq-12|vq \geq 12] - c_2 & \text{if visiting micro-stock site directly} \\ U^{\mathrm{O}}(v) = 0 & \text{if taking the outside option} \end{cases}, \tag{1}$$

where $m(v) = \max\{U^{\mathrm{M}}(v) - c_3, 0\}$ indicates the continuation value of not purchasing the premium image—the user can still choose whether to take the outside option or to search for a replacement image on the micro-stock site with a total search cost of $c_2 + c_3$. (c_2 is included in $U^{\mathrm{M}}(v)$, which we will explain later.) In words, the expected payoff from first visiting the premium site, $U^{\mathrm{P}}(v)$, is the probability that the actual price is sufficiently low that the net payoff from purchasing a premium license ($v-p$) is greater than the continuation value ($m(v)$), multiplied by the conditional expected net payoff, plus the probability that the actual price is too high, multiplied by $m(v)$, minus the cost of searching for the price information. If going to the micro-stock site directly, the user's expected payoff, $U^{\mathrm{M}}(v)$, is the probability that she will find an image that yields her a payoff higher than \$12, multiplied by the conditional expected net payoff, minus the search cost c_2.

Let $\overline{v}$ be the user who is indifferent between going to the micro-stock site directly and visiting the premium site first (that is, the solution to $U^{\mathrm{M}}(v) = U^{\mathrm{P}}(v)$) and $\underline{v}$ be the user who is indifferent between visiting the micro-stock site directly and taking the outside option (that is, the solution to $U^{\mathrm{M}}(v) = 0$). Furthermore, denote $\overline{\overline{v}} = \tilde{p} + m(v)$ as the threshold user who finds it worthwhile to purchase the premium

[23] In principle, the user may still visit the premium site after visiting the micro-stock site first. Empirically, the data show that for users who searched both sites, 90 percent started with the premium site. This is intuitive because the premium option is listed before the micro-stock option in the email. Theoretically, with additional technical assumptions, the threshold result described in Proposition 1 still holds. The predictions change only slightly and in ways that do not alter our interpretation of the data. Given that abstracting away from this possibility is not too concerning either empirically or theoretically (at least in the context of our current conceptual framework), for simplicity, we choose not to explicitly model it.

image after discovering the actual premium price, $\tilde{p}$; and $\underline{v}'$ as users who are indifferent between continuing to search the micro-stock site and taking the outside option after spending time on the premium site (that is, the solution of $U^{\mathrm{M}}(v) - c_3 = 0$).

To simplify the analysis, we make two technical assumptions (which are explained in Appendix C.1) that exclude scenarios that are not relevant to our empirical context. Assumption 1 guarantees that each of the three initial search choices is optimal for at least some users, and Assumption 2 assumes that the actual premium price is sufficiently high that only a small percentage of users who visit the premium site find it worthwhile to purchase the premium image (that is, $\bar{\bar{v}} > \bar{v}$).

Under Assumption 1, the following result shows that the user follows a threshold rule for the three initial search options (see all the proofs in the Appendix):[24]

Proposition 1. *There exist unique $\underline{v} < \bar{v}$ such that the user would take the outside option if $v < \underline{v}$; go to the micro-stock site directly if $\underline{v} < v < \bar{v}$; and go to the premium site first if $v > \bar{v}$.*

With this simple setup, we derive the probability of users searching either of the two licensing sites. The probability of users visiting the premium site is simply:

$$\mathrm{S}^{\mathrm{P}} = \Pr(v > \bar{v}). \tag{2}$$

The total probability of users searching the micro-stock site is the sum of the probabilities of (i) those who search the micro-stock site directly (i.e., $\Pr(\underline{v} \leq v < \bar{v})$); and (ii) those who search the micro-stock site after visiting the premium site, which is calculated differently in two different scenarios. Note that we can reduce our discussions to these two scenarios, because $\underline{v} < \bar{v} < \bar{\bar{v}}$ under Assumptions 1 and 2.

Figure 2a illustrates the first scenario, in which $\bar{v} > \underline{v}'$. In this case, all users who first visit the premium site and discover that the price is too high will find it worthwhile to search the micro-stock site afterwards. This happens when the additional search cost for a replacement image c_3 is relatively small, including the benchmark case in which $c_3 = 0$ and, consequently, $\underline{v}' = \underline{v}$.

Figure 2b illustrates the second scenario, in which $\bar{v} < \underline{v}'$. In this case, at least some users, after spending time on the premium site, find it too costly to continue to search the micro-stock site (and, hence, take the outside option instead). This scenario happens when the additional search cost for a replacement image c_3

[24] For group "Micro," we see some support for this statement: the mean of log(Total stock images on site+1), a noisy proxy for a firm's demand for images, is 1.68 for users who logged on to at least one of the two sites and 1.37 for users who did neither (p-value is 0.05).

is sufficiently high.[25] The probability of users visiting the micro-stock site is, thus:

$$\mathrm{S^M} = \begin{cases} \Pr(\underline{v} \leq v \leq \overline{\overline{v}}) & \text{if } \overline{v} \geq \underline{v}' \\ \Pr(\underline{v} \leq v < \overline{v}) + \Pr(\underline{v}' \leq v \leq \overline{\overline{v}}) & \text{if } \overline{v} < \underline{v}' \end{cases} . \tag{3}$$

Notice that, relative to the first scenario, there is a gap in the second scenario in the users ($\overline{v} < v < \underline{v}'$) whose willingness is quite high but who end up not searching for a replacement image. Thus, the second scenario highlights the potential downside of first visiting the premium site relative to searching the micro-stock site directly—i.e., the option of searching for a replacement image becomes less attractive after users spend time on the premium site because they face a higher marginal search cost—and the potential welfare loss from these users with a relatively high willingness to pay.

In the following, we present the effects of the two interventions on the likelihood of searching both licensing sites, $\mathrm{S^P}$ and $\mathrm{S^M}$. For each intervention, we first present the model's predictions and then the empirical results.

3.3.1 Effects of image recommendations on the likelihood of searching both sites

In the context of our model, image recommendations may affect the expected payoff from searching the micro-stock site (either directly or after visiting the premium site) through two mechanisms: (i) reducing the direct costs of searching for a replacement image; and (ii) revealing information about the quality of these cheaper images, which may increase or decrease the user's expected benefit from searching the site, depending on how the revealed quality compares to the user's prior belief. Prediction 1 considers these two mechanisms separately.

Prediction 1.

i) Suppose that image recommendations reduce the direct costs of searching for a replacement image (that is, by lowering c_2 and c_3).

a) If $c_3 = 0$ to start with, S^M will increase and S^P will decrease; but

b) if c_3 is sufficiently large, it is possible that both S^M and S^P will increase.

ii) Suppose that image recommendations provide more-precise information about image quality.

a) If the revealed quality is higher than the user's prior beliefs, S^M will increase and S^P will decrease;

b) otherwise, S^M will decrease and S^P will increase.

[25] See Corollary 1 in Appendix C.1 for a sufficient condition for the existence of the second scenario. Note that this scenario includes the case of $\underline{v}' > \overline{\overline{v}}$, which we do not illustrate in Figure 2b. But in this case, the second segment of the probability of visiting the micro-stock segment as is described in the equation below, $\Pr(\underline{v} \leq v < \overline{v}) + \Pr(\underline{v}' \leq v \leq \overline{\overline{v}})$, will simply be reduced to zero.

Prediction 1.i derives the effects of image recommendations, taking into account only the first mechanism of reducing the direct search costs. We consider two scenarios. In the benchmark case of $c_3 = 0$ (illustrated in Figure 2a), Prediction 1.i.a shows that image recommendations should increase the total likelihood of visiting the micro-stock site but decrease the likelihood of visiting the premium site. Intuitively, making the process of searching for a replacement image easier has a "cannibalization effect" that decreases the appeal of searching for the premium price (that is, $\bar{v}$ in Figure 2a will increase).

However, if c_3 is sufficiently large (consider the scenario illustrated in Figure 2b), Prediction 1.i.b shows that it is possible that the likelihood of search increases on both sites. The intuition is that having a specific replacement image to fall back on also reduces this *extra* cost of searching the micro-stock site after first visiting the premium site. Note that this additional effect of image recommendations increases the expected payoff from first visiting the premium site ($U^{\mathrm{P}}(v)$) and does not influence the expected payoff from going to the micro-stock site directly ($U^{\mathrm{M}}(v)$). Thus, when c_3 is sufficiently large, this additional effect is strong enough to offset the cannibalization effect, and we should observe an increase in the likelihood of visiting the premium site, as well as the micro-stock site.[26]

Prediction 1.ii derives the effects of this intervention via the second mechanism of revealing image quality. It shows that the effects depend on how the revealed image quality compares to users' expectations.[27] With this mechanism, however, because there is no additional effect to offset the cannibalization effect discussed above, we cannot observe an increased likelihood of search on both sites.

Table 5 presents the OLS regression results on the likelihood of clicking onto either the premium site or the micro-stock site. Column 1 shows that users in our experiment are significantly more likely to search the premium site when they receive image recommendations than when they do not. The difference is 1.7 percentage points, or a 23-percent increase relative to the two groups without image recommendations (p-value is 0.029).[28] At the same time, column 2 shows that users who receive image recommendations are also more likely than those without recommendations to search the micro-stock site: the difference is 5.5 percentage points or a 96-percent increase (p-value < 0.001). Conditional on visiting the micro-stock site, however, unreported regression results show that users given recommendations are, on average, 8.8 percentage points (or 25-percent) less likely to license than users not given recommendations (p-value

[26] Note that even in the benchmark case considered in Prediction 1.i.a, the *standalone* payoff from first visiting the premium site, $U^{\mathrm{P}}(v)$, also increases with image recommendations. This is because making the subsequent search on the micro-stock site easier improves the user's continuation value of not purchasing the premium image. But this increase will be smaller than the increase in the expected payoff from searching the micro-stock site directly, thus resulting in a smaller likelihood of searching the premium site (and, hence, the cannibalization effect).

[27] In the appendix, we show that these predictions do not depend on the value of c_3.

[28] A further breakdown is intuitive: column 3 in table 5 shows that when the premium-image price information is not given, the difference in the probability of searching the premium site is 2.2 percentage points (p-value is 0.093), which is larger than the 1.1 percentage point difference when users are given the premium-price information (p-value 0.146). Note that the interaction term shows that the difference between these two effects is not statistically significant.

0.095).

The fact that the search likelihood increases for both the premium and the micro-stock sites is consistent with the idea that users face non-trivial and increasing search costs, and that image recommendations reduce these costs (Prediction 1.i.b). This not only attracts users away from the outside option, but also mitigates the opportunity cost of first learning the premium price.[29] The second mechanism of revealing image quality may contribute to the increase in the likelihood of searching the micro-stock site—for example, if users' prior beliefs about image quality are sufficiently pessimistic and images recommendations correct this belief. However, this mechanism cannot explain the increased likelihood of search on both sites.

Users who respond to image recommendations include those in the ranges of $v < \underline{v}$ and $\overline{v} < v < \underline{v}'$ in Figure 2b who do not find it worthwhile searching for a replacement image without image recommendations, but do with recommendations. Marginal users (those with $v < \underline{v}$) with relatively lower willingness to pay are drawn away from taking the outside option directly. This may help explain the lower licensing rate, conditional on browsing the micro-stock site, for users with recommendations relative to those without.

We also find that, conditional on licensing, half of the images purchased by users who received recommendations are recommended images. In comparison, five percent of the users who do not receive recommendations, but who license an image found through independent search, ultimately license images that coincide with images generated by the same algorithm. This large difference (45 percentage points) confirms that image recommendations influence users' choices. This large difference also provides an interesting contrast to Sharma et al. (2015), who find that in the case of Amazon, at least 75 percent of recommendation click-throughs would have happened even without Amazon's recommender system (in our experiment, this number is ten percent).[30] One potential explanation of the difference between the two results is that Amazon offers several additional sources of information about alternative products, such as reviews, that may lower the marginal value of product recommendation on a standalone basis.

Using a 'similarity' measure provided by the Agency, we find that, conditional on licensing, the average 'similarity' between the licensed image and the infringed image is statistically similar for users who did and did not receive image recommendations. This result is consistent with Horton (2017), who finds that a significant fraction of employers using oDesk, a large online labor market, follow algorithmic recommendations when extending offers to potential recruits, but that the workers recruited through recommendation are observationally similar to workers that employers would have recruited on their own in the absence of any

[29] In our experimental sample, the additional visits to the premium sites are 'wasted,' because the premium price is too high for almost all infringers. If, however, the price of the infringed products is affordable for many users with $v < \overline{v}$ (that is, Assumption 2 is violated), many of these additional visits will be converted to purchases.

[30] The ten-percent number is derived from five percent (of licensed images by users not given recommendations coinciding with images generated by the algorithm) divided by 50 percent (of licensed images by users receiving recommendations being recommended images).

algorithm recommendations. One potential explanation for this type of result is the selection effect discussed above; that is, users who license, or recruiters who make offers despite the absence of a recommendation, are likely to have a greater willingness to pay or a lower search cost. Another possible explanation is that there are potentially many images or workers that are close substitutes for each other; thus, algorithms may work mainly through providing convenience rather than superior match quality over independent search.

3.3.2 Effects of premium-image price information on the likelihood of searching both sites

The following prediction describes the effects of providing the premium price information in the email:

Prediction 2.

i) S^P will decrease.

ii) If $c_3 = 0$ to start with, S^M does not change; if c_3 is sufficiently high, S^M may increase.

Recall that under Assumption 2, the premium price is sufficiently high that in group Micro, only a small percentage of users who visit the premium site find it worthwhile to purchase the premium image (that is, $\overline{\overline{v}} > \overline{v}$). With the price range revealed in the email, only users who find it worthwhile to purchase the premium image will visit the premium site. Moreover, without having to spend time on the premium site to find out the price, users also face a lower search cost on the micro-stock site (that is, c_2 instead of $c_2 + c_3$), making the micro-stock option more attractive.[31] For both reasons, the probability of visiting the premium site will decrease with the provision of the premium-price information.

Prediction 2.ii shows that the way in which the premium-price information affects users' likelihood of searching the micro-stock site depends, again, on how much more costly it is to search for a replacement image after visiting the premium site. In the benchmark case of $c_3 = 0$ (Figure 2a), the total likelihood of searching the micro-stock site should stay the same because neither $\underline{v}$ nor $\overline{\overline{v}}$ will be affected by knowing the premium price. However, if c_3 is sufficiently large (Figure 2b), we may observe more searches on the micro-stock site by price-informed users because these users—without having to waste time discovering the premium-price information—now face a lower search cost on the micro-stock site (in particular, users with $\overline{v} \leq v \leq \underline{v}'$ will now continue to search the micro-stock site rather than taking the outside option).

Column 1 of table 5 shows that, within the four treatment groups, the likelihood of searching the premium site after receiving the price information is, on average, 8.8 percentage points (or 69 percent) lower than that for the two groups that do not receive this information (p-value < 0.001). This large decrease is consistent with the simple comparison between the two control groups with only the premium-image licensing option; between these two groups, the price information reduces the likelihood of visiting the premium

[31] With the price information revealed, the user who is indifferent between visiting the premium site and continuing to search the micro-stock site is $\tilde{p} + \max\{U^{\mathrm{M}}(v), 0\}$, which is greater than or equal to $\overline{\overline{v}} = \tilde{p} + \max\{U^{\mathrm{M}}(v) - c_3, 0\}$.

site from 0.19 to 0.045 (p-value < 0.001). These results suggest that the actual premium price is substantially higher than users' expectation such that many users, given this information, no longer consider the possibility of purchasing the premium image.

Consistent with an increasing marginal search cost (the second scenario considered in Prediction 2.ii), column 2 of table 5 shows that providing premium-price information increases the likelihood of visiting the micro-stock site by 1.6 percentage points (or 15 percent, p-value 0.042). Unlike the case of image recommendations, we do not observe a lower likelihood of licensing conditional on searching the micro-stock site.[32] This is consistent with the result that those induced to search the micro-stock site after receiving the premium-price information are all infra-marginal users (that is, $\overline{v} \leq v \leq \underline{v}'$ in Figure 2b), who would have searched for the premium price without this intervention.

Note that users who respond to the price information are a subset of those who respond to image recommendations. For these common 'compliers,' if search frictions are already reduced by one intervention, the incremental effect of the other is smaller. This is consistent with the negative interaction term of the two interventions on the likelihood of visiting the micro-stock site (column 4 of table 5). In particular, the effect of image recommendations is 6.9 percentage points for users not given the premium-price information (or a 186-percent increase relative to group "Micro"). For users given the price information, the effect of image recommendations drops to 4.0 percentage points (or a 52-percent increase relative to group "Micro+Price" and p-value < 0.001). The effect of the premium-price information is 3.6 percentage points for users not given image recommendations (or 97-percent increase relative to group "Micro"), but it mostly disappears for users given recommendations (0.7 percentage points and p-value is 0.51).

Our results also show that when starting from the same baseline (group "Micro"), the effect of image recommendation is economically larger than that of premium-price information (186 versus 97 percent). This is intuitive, as the former reduces the cost of finding a replacement image directly, whereas the latter reduces the cost only indirectly, by guiding people to the affordable option faster. However, the overall effect of the premium-price information on the eventual licensing probability is similar to that associated with image recommendations (see the two coefficients 0.224 and 0.248 in column 4 of table 4). This is because many users affected by image recommendations are marginal users, while those that respond to price information are all infra-marginal users.

3.3.3 Alternative explanations

Our model focuses on a reduction in search frictions. There might be alternative explanations for the effects of these interventions, and we briefly discuss some of them below.

[32] Unreported regression results show that conditional on browsing the micro-stock site, the coefficient of 'With premium-image price information' is 1.3 percentage points and has a p-value is 0.732.

First, the premium-price information may increase the search likelihood of the micro-stock site if it makes the micro-stock option look more attractive. For example, seeing that the premium price is significantly greater than the expected level (which serves as a reference point), the user may perceive a greater (standalone) utility from searching for a micro-stock image.[33] Consider the comparison between the two treatment groups that are given image recommendations. For these two groups, the search frictions for replacement images are already low, and, thus, the friction-reduction channel of the premium-price information is relatively small. The results show that the extra effect of the premium-price information is very small and statistically insignificant for both the likelihood of browsing the micro-stock site and the licensing of micro-stock images. This suggests that these price effects, though plausible, may have limited standalone explanatory power in this particular context.

Second, even though the general tone of the emails is friendly, the price information may increase the perceived likelihood that the Agency will escalate its enforcement because the price is higher than expected for most firms (Luo and Mortimer (2017)). This may result in higher licensing rates. We do not find supporting evidence for this conjecture in the longer-term outcome data that we collected for a random subsample of our sample firms. We find that, on average, groups given the price information are about as likely as groups not given this information to continue displaying the infringed premium image on their websites (0.127 versus 0.123, and p-value is 0.902).[34]

Third, image recommendations may simply grab people's attention because of the salient visual effect of the colorful images.[35] Many users may click onto the micro-stock site only to see larger versions of these images, rather than looking for a replacement image. This may explain why groups with image recommendations click onto the micro-stock site more often and may also explain the lower licensing rate conditional on browsing. However, these explanations cannot explain why the overall licensing rate is significantly higher among users that receive image recommendations or why users tend to follow recommendations when licensing images.

[33] Thaler (1985) develops a model of consumer behavior that incorporates transaction utility (the perceived merits of the deal). An implication of the model is that an actual transaction price that is lower than the consumer's reference price provides positive transaction utility and, thus, results in greater consumption. He uses the model to rationalize phenomena such as why we often observe that manufacturers' suggested retail prices are much higher than market prices. Note that Thaler (1985) considers only one product, whereas there are two competing products in our setting. A direct application of Thaler (1985) to the premium price intervention in our context will lower the (standalone) payoff from searching the premium site. But the (standalone) payoff from searching the micro-stock site may also increase if a significantly higher-than-expected price for the competing product gives users a positive 'transaction utility' from choosing the micro-stock option.

[34] See Appendix C for a detailed description of this dataset and our take-aways.

[35] Visual salience is the "distinct subjective perceptual quality which makes some items in the world stand out from their neighbors and immediately grab our attention" (Laurent Itti, 2007, Scholarpedia). Milosavljevic et al. (2012), for example, show that at rapid decision speeds, visual saliency may influence choices more than preferences do.

4 Discussion

Taken together, our findings provide evidence of the following features in the digital images market: (i) users face search and information frictions in the ex-ante licensing market; (ii) infringing use can reveal information about overall consumer demand and preferences for specific products that may be costly to obtain ex-ante; and (iii) simplifying the search and transaction process is important for small-value transactions. All three features may characterize copyright markets more generally. For example, infringing use reveals information about underlying demand conditions in many settings; and significant search and other transactional frictions are often identified in the music industry with respect to use for follow-on productions.

The effectiveness of an ex-post monetization approach is likely to be context-dependent. In particular, it should depend on the value created by having an ex-post market and the costs of monitoring infringing use. Our results suggest that ex-post monetization potentially expands the market to include consumers who would not have been active in an ex-ante market due to high search and transactions costs, and it can take advantage of the demand information revealed by infringing use. Thus, roughly speaking, the larger are search and transactions frictions in the ex-ante market, and the more difficult demand is to predict ex-ante, the larger is the potential value of an ex-post monetization approach.

Whether it is worthwhile to develop an ex-post market depends also on the cost of monitoring infringing use.[36] When fixed costs of monitoring are high (e.g., the costs of developing and refining machine learning algorithms), scale will be important. Thus, developing monitoring technologies in-house may make sense only for the largest firms and platforms (such as YouTube and Facebook) or for industry-wide third-party providers.[37] Given the increasing importance of managing online use in general, some of these fixed investments may also be shared with the ex-ante markets. When marginal costs are non-trivial, selective monitoring—such as monitoring only those assets that are most frequently purchased (and, hence, more likely to be infringed upon) or only infringers likely to have a high willingness to pay—may be more cost-effective than universal monitoring. In some circumstances, crowd-sourcing detection may tap into creators' incentives to protect their own products or employees' incentives to benefit from a sales lead.

Finally, two conditions are taken as given in our setting but may be important for fast ex-post resolution. First, infringement use can be precisely identified and is clear to all relevant parties. In other settings,

[36] Given that our experiment is a tiny part of the Agency's compliance operation at the time, it is hard to properly allocate costs. That said, simple back-of-envelope calculations show that if we stick to the parameters of our experiments, this approach is unlikely to be profitable. As discussed later in this section, we expect the cost-benefit comparison to change significantly once we consider: (i) other products of the Agency that are licensed (and infringed) in substantially greater scope and for which users can simply pay without having to find replacement images; (ii) larger infringing firms with greater willingness to pay; and (iii) potentially more cost-effective monitoring technologies.

[37] Digimarc and Thomson Reuters provide infringement-detection services for movies, music, software, and publishing. As we discuss below, YouTube has invested $100m since 2007 in an auto-detection technology of use of audio and video content.

ambiguity over whether infringement has occurred may add substantial frictions and deter fast resolution. For example, while infringement is relatively clear for copyrighted goods based on visual examination and fingerprinting technologies using meta data, this is often not the case for patented technologies. Within copyright settings, use occasions that border on fair use may also lead to confusion. Second, rights holders' ability to commit to a pricing level or to a revenue-sharing regime that is generally considered 'fair' seems critical for removing costly ex-post negotiations, including hold-up, and for preventing static maximization of enforcement revenues from becoming the primary goal.

In light of the above discussion, we turn to some concrete examples of how the idea of ex-post monetization may apply to contexts beyond our experimental setting. Our experimental setting is one in which making users aware of more-affordable alternative options and reducing their costs of finding a replacement product are critical for inducing ex-post licensing. More generally, rights holders may employ similar approaches if they own multiple tiers of products that are substitutable but priced at different levels, and infringers are unlikely to purchase the infringed products. Potential markets sharing these characteristics include various software products and textbooks for which rights holders tend to version their products and price them discriminately (Shapiro and Varian (1998)).

In contrast, if infringers are likely to purchase the infringed products, rights holders may be better off focusing on inducing same-product licensing—for example, by providing convenient price information and making payment easier. Users in these situations may be more willing to license the infringed product because they can avoid the costs associated with using a replacement product (e.g., loss in fit, search costs for replacement products, and additional costs of replacing the existing product with another). Within the digital image market, for example, rights-managed premium images (the type of infringed products our experiment focuses on) account for less than one percent of the Agency's total portfolio. The volume of unauthorized use of the Agency's other products—micro-stock images and royalty-free premium images (which cost between \$50 and \$500 but can be used without any restrictions)—is vastly greater and is expected to continue to increase, given the shrinking share of rights-managed premium licenses in this industry. If one were to consider infringing use of these other products by a more representative sample of infringing firms, our conjecture is that both the likelihood of licensing and the expected revenue per case would be substantially higher than the levels in our experiment. If the infringed product is already the cheapest option, the natural response is to provide only the option of licensing the same image and to make the transaction process easier. We are aware of a different internal trial on a different population of firms by the Agency before our experiment, which shows that providing a simple direct payment link by email leads to more licensing. If the infringed product is at an intermediate price level, the Agency needs to also consider whether to recommend similar but cheaper options along with the option of licensing the focal product. This choice likely depends on the Agency's belief about the users' willingness to pay and the substitutability of the products.[38]

Lastly, we consider situations in which replacement is not necessary, but rights holders may monetize indirectly (e.g., via ad placement). An example is YouTube's Content ID program, which automatically detects infringing use of audio and visual reference files (contributed by rights holders) when user-generated videos are uploaded. Once infringing use is confirmed, YouTube applies one of three options, which is chosen *ex-ante* by rights holders: (i) block usage; (ii) track viewing statistics; or (iii) monetize through ad placement.[39] The YouTube Content ID program may represent a best-case scenario for ex-post monetization, in the sense that it (i) has the scale to spread fixed investments in detection technologies; (ii) takes advantage of the platform's existing infrastructures for ad placement and payment, significantly lowering the various types of transactions costs, such as those associated with measuring revenues and enforcing and processing payment; and (iii) leverages a higher willingness to pay by advertisers. Finally, as Gans (2015) points out, by committing rights holders to a pre-determined revenue-sharing regime, this program reduces significant frictions in ex-post bargaining (e.g., due to the parties' inability to agree on the added value of the copyrighted inputs).

5 Conclusion

In this study, we explore the feasibility of an ex-post approach to monetizing copyrighted products. We contacted infringing firms of expensive digital images and directed them toward a significantly cheaper, replacement product. We further designed two interventions to reduce search costs for replacement products and for the price comparison between the two products. We find that awareness of the cheaper option leads to a substantial increase in the licensing probability. Furthermore, consistent with the idea that infringing use is informative of demand, the licensing rate of the cheaper option by our treatment groups is several times greater than the most effective email marketing campaigns at the Agency for the same type of products. Both interventions that intend to reduce search frictions have a large positive effect on the probability of search and, eventually, licensing. Our results suggest that ex-post licensing may expand the market to include consumers who would not have been active in an ex-ante market, and that rights holders can create value in the ex-post market by minimizing search and transactions costs.

[38] We do not examine the impacts on premium purchases in our experiment because they are rare. If the infringed products are, in fact, moderately priced, our model predicts that both interventions—provision of the price of the infringed products and recommending similar but cheaper products—would increase the likelihood of same-product purchases if the actual price is lower than users' expectations. Relative to price-information provision, recommending cheaper products is likely to generate a smaller positive effect on the likelihood of same-product purchases. This is partly because these recommendations make the cheaper option more attractive, cannibalizing sales of a higher-priced product. This effect suggests that if rights holders are reasonably confident that the infringers can afford the medium-priced products, they are likely to be better off *not* recommending a cheaper product.

[39] According to Google's recent updates, it has invested about $100m in the algorithm since 2007; rights holders choose to monetize in 90 percent of the cases and have collectively received about $3B in ad revenues. Source: "How Google Fights Piracy," by Google, November 2018, pages 24-27.

We conclude by cautioning that our results do not reflect potential equilibrium effects. In equilibrium, rights holders should respond to substitution between premium and cheaper offerings, as well as between the ex-ante and the ex-post markets.[40] Customers in the ex-ante market may substitute to the ex-post market if this makes them better off, given the relative prices and probability of detection. Similarly, easier licensing options for cheaper products may cannibalize sales of premium products, leading to price changes for both products. Beyond any pricing responses, rights holders may also change the search and transaction process of the ex-ante market, and competitors may respond with new pricing and product offerings. Theoretical and empirical research on the co-existence of ex-post and ex-ante markets and their equilibrium effects would be interesting and highly relevant for policy makers and managers.

References

Adomavicius, G. and A. Tuzhilin (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering 17*, 734–749.

Aguiar, L. and J. Waldfogel (2018). As streaming reaches flood stage, does it stimulate or depress music sales? *International Journal of Industrial Organization 57*, 278–307.

Alba, J., J. Lynch, B. Weitz, C. Janiszewski, R. Lutz, A. Sawyer, and S. Wood (1997). Interactive home shopping: Incentives for consumers, retailers, and manufacturers to participate in electronic marketplaces. *Journal of Marketing 61*, 38–53.

Anand, B. N. and R. Shachar (2011). Advertising, the matchmaker. *RAND Journal of Economics 42*, 205–245.

Avery, C., P. Resnick, and R. Zeckhauser (1999). The market for evaluations. *American Economic Review 89*, 564–584.

Bakos, J. Y. (1997). Reducing buyer search costs: Implications for electronic marketplaces. *Management Science 43*, 1613–1630.

Bakos, Y., E. Brynjolfssson, and D. Lichtman (1999). Shared information goods. *Journal of Law and Economics XLII*, 117–155.

Belleflamme, P. and M. Peitz (2012). Digital piracy: theory. In M. Peitz and J. Waldfogel (Eds.), *The Oxford Handbook of the Digital Economy*, pp. 489–530. Oxford: Oxford University Press.

[40] To study equilibrium effects, one needs a dynamic model in which users decide whether to search and, upon searching, whether to purchase a given option in the ex-ante market, or to pirate (e.g., using a product found through a search engine) and pay upon being caught in the ex-post market. Hua and Spier (2019) is the only paper we are aware of that studies software licensing with and without an ex-post market. In their model, there are no search costs and there is a single product. A key result of their paper is that the presence of an ex-post market facilitates price discrimination—high-type consumers purchase in the ex-ante market, whereas low-type consumers pirate and pay in the ex-post market upon being caught. The situation becomes more complicated with multiple products. Moreover, in our conceptual framework, apart from willingness to pay, users may also face different search costs. Thus, the presence of an ex-post market may also segment users by their search costs; for example, buyers or buying situations with low search costs tend to transact in the ex-ante market, whereas those with higher search costs transact in the ex-post market. The impact of this additional effect, alone, on equilibrium prices likely depends on how consumers' willingness to pay correlates with search costs.

Belleflamme, P. and M. Peitz (2014). Digital piracy: An update. ECORE Discussion Paper.

Belluf, T., L. Xavier, and R. G. . . I. P. RecSys. (2012). Case study on the business value impact of personalized recommendations on a large online retailer. In *Proceedings of the sixth ACM conference on recommender systems*, pp. 277–280.

Bessen, J. E. and M. J. Meurer (2006, May). Patent Litigation with Endogenous Disputes. *American Economic Review 96*(2), 77–81.

Bessen, S. and S. Kirby (1989). Private copying, appropriability, and optimal copying royalties. *Journal of Law and Economics 32*, 225–280.

Bhattacharjee, S., R. D. Gopal, K. Lertwachara, and J. Marsden (2006). Impact of legal threats on individual behavior: An analysis of music industry actions and online music sharing. *Journal of Law and Economics 49*, 91–114.

Brynjolfsson, E. and M. Smith (2000). Frictionless commerce? a comparison of internet and conventional retailers. *Marketing Science 46*, 563–585.

Carmi, E., G. Oestreicher-Singer, U. Stettner, and A. Sundararajan (2017). Is oprah contagious? the depth of diffusion of demand shocks in a product network. *MIS Quarterly 41*, 207–221.

Choi, J. P. and H. Gerlach (2018). A model of patent trolls. *International Economic Review 59*, 2075–2106.

Cohen, L., U. Gurun, and S. D. Kominers (2014, August). Patent trolls: Evidence from targeted firms. Management Science (forthcoming).

Conner, K. R. and R. P. Rumelt (1991). Software piracy: an analysis of protection strategies. *Management Science 37*, 125–139.

Danaher, B., S. Dhanasobhon, M. D. Smith, and R. Telang (2010). Converting pirates without cannibalizing purchasers: the impact of digital distribution on physical sales and internet piracy. *Marketing Science 29*, 1138–51.

Danaher, B. and M. Smith (2014). Gone in 60 seconds: The impact of the megaupload shutdown on movie sales. forthcoming International Journal of Industrial Organization.

Danaher, B., M. Smith, R. Telang, and S. Chen (2014). The effect of graduated response anti-piracy laws on music sales: Evidence from a natural experiment in france. *Journal of Industrial Economics 62*, 541–553.

Fleder, D. and K. Hosanagar (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science 55*, 697–712.

Fradkin, A. (2017). Search, matching, and the role of digital marketplace design in enabling trade: Evidence from airbnb. Working paper.

Gans, J. (2012). *Information Wants to be Shared*. Boston, MA: Harvard Business Review Press.

Gans, J. (2015). Remix rights and negotiations over the use of copy-protected works. *International Journal of Industrial Organization 41*, 76–83.

Glückler, J. and R. Panitz (2013). Survey of the global stock image market 2012. part i: Players, products, business. Technical report, Heidelberg: GSIM Research Group.

Godinho de Matos, M., P. Ferreira, and M. D. Smith (2018). The effect of subscription video-on-demand on piracy: Evidence from a household-level randomized experiment. *Management Science 64*, 5461–5959. Management Science (forthcoming).

Horton, J. (2017). The effects of algorithmic labor market recommendations: Evidence from a field experiment. *Journal of Labor Economics 35*, 345–385.

Horton, J. (2019). Buyer uncertainty about seller capacity: Causes, consequences, and a partial solution. *Management Science 65*, 3449–3947.

Hua, X. and K. Spier (2019). Settling lawsuits with pirates. Work-in-progress.

Jannach, D. and K. Hegelich (2009). A case study on the effectiveness of recommendations in the mobile internet. *Proceedings of the third ACM conference on Recommender systems*, 205–208.

Kretschmer, T. and C. Peukert (2017). Video killed the radio star? online music videos and recorded music sales. Working paper.

Lewis, R. A. and J. M. Rao (2015). The unfavorable economics of measuring the returns to advertising. *Quarterly Journal of Economics 130*, 1941–1973.

Lu, J., D. Wu, M. Mao, W. Wang, and G. Zhang (2015). Recommender system application developments: A survey. *Decision Support Systems 74*, 12–32.

Luo, H. and J. H. Mortimer (2017). Copyright enforcement: Evidence from two field experiments. *Journal of Economics and Management Strategy 26*, 499–528.

Lynch, J. G. and D. Ariely (2000). Wine online: Search costs affect competition on price, quality, and distribution. *Marketing Science 19*, 83–103.

Menell, P. S. (2014). This american copyright life: Reflections on re-equilibrating copyright for the internet age. UC Berkeley Public Law Research Paper No. 2347674.

Menell, P. S. (2016). Adapting copyright for the mashup generation. *University of Pennsylvania Law Review 164*, 441–512.

Milosavljevic, M., V. Navalpakkam, C. Koch, and A. Rangel (2012). Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology 22*, 67–74.

Oberholzer-Gee, F. and K. Strumpf (2007). The effect of file sharing on record sales: An empirical analysis. *Journal of Political Economy 115*, 1–42.

Peitz, M. and P. Waelbroeck (2006a). Piracy of digital products: a critical review of the theoretical literature. *Information Economics and Policy 18*, 449–76.

Peitz, M. and P. Waelbroeck (2006b). Why the music industry may gain from free downloading: the role of sampling. *International Journal of Industrial Organization 24*, 907–13.

Reimers, I. (2016). Can private copyright protection be effective? evidence from book publishing. *Journal of Law and Economics 59*, 411–440.

Resnick, P. and H. R. Varian (1997). Recommender systems. *Communications of the ACM 40*, 56–58.

Rob, R. and J. Waldfogel (2006). Piracy on the high c's: Music downloading, sales displacement, and social welfare in a sample of college students. *Journal of Law and Economics 49*, 29–62.

Scott Morton, F. and C. Shapiro (2014). Strategic patent acquisitions. *Antitrust Law Journal 79*, 463–499.

Shapiro, C. (2010). Injunctions, hold-up, and patent royalties. *American Law and Economics Review 12*, 509–557.

Shapiro, C. and H. Varian (1998). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press.

Sharma, A., J. M. Hofman, and D. J. Watts (2015). Estimating the causal impact of recommendation systems from observational data. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, New York, NY, USA, pp. 453–470. ACM.

Shy, O. and J.-F. Thisse (1999). A strategic approach to software protection. *Journal of Economics & Management 8*, 163–190.

Smith, M. D. and R. Telang (2009). Competing with free: the impact of movie broadcasts on dvd sales and internet piracy. *MIS Quarterly 33*, 321–38.

Smith, M. D. and R. Telang (2016). *Streaming, Sharing, Stealing: Big Data and the Future of Entertainment*. The MIT Press.

Tadelis, S. and F. Zettelmeyer (2015). Information disclosure as a matching mechanism: Theory and evidence from a field experiment. *American Economic Review 105*, 886–905.

Takeyama, L. N. (1994). The welfare implications of unauthorized reproduction of intellectual property in the presence of demand network externalitie. *Journal of Industrial Economics 42*, 155–167.

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science 4*, 199–214.

Varian, H. R. (2005). Copying and copyright. *The Journal of Economic Perspectives 19*, 121–138.

Waldfogel, J. (2012). Digital piracy: empirics. In M. Peitz and J. Waldfogel (Eds.), *The Oxford Handbook of the Digital Economy*, pp. 512–46. Oxford University Press.

Waldfogel, J. (2018). How digitization has created a golden age of music, movies, books, and television. *Journal of Economic Perspectives*.

Zhang, L. (2016). Intellectual property strategy and the long tail: Evidence from the recorded music industry. *Management Science 64*, 24–42.

Figure 1: Template illustration—emails to group "Micro + Rec"

SUBJECT LINE: Unauthorized image use on your site
PREHEADER: Action required for resolution

EMAIL:
Dear *[COMPANY NAME]*,

We are contacting you because the below imagery represented exclusively by [redacted], a global digital media provider, is being displayed on your company's website [BUSINESS URL]:
[RM THUMBNAIL IMAGE] [RM THUMBNAIL SCREEN SHOT]

We are thrilled you like our imagery! However, we have no record of an active license for its commercial use on your site. We understand the unlicensed use may be accidental, but in fairness to our photographers, we ask that you take one of the following actions:

1. **Provide proof of a license that covers your use**
   If you do have a license, or if you believe you have mistakenly received this email, please email your licensing or other relevant information to [redacted], and indicate the reference number [Usage ID] in your e-mail.

2. **Purchase a license here** [RM IMAGE ADP PAGE] **to cover future use**
   This image is only available under a premium Rights Managed license.

3. **Replace the premium image with a high-quality, but more affordable, image from** [redacted]
   Our [redacted] site offers a large collection of images that can be used for commercial websites with costs as low as $12 per image. We have selected the images below for you as possible replacements:
   [IMAGE URL SIMILAR 1] [IMAGE URL SIMILAR 2] [IMAGE URL SIMILAR 3] [IMAGE URL SIMILAR 4]

In order for us to associate your purchase with this incident, please use the email address to which this notice was sent to create an account, and enter the reference number [Usage ID] in the "Purchase order number" field (click "Add note") when placing your order.

Please understand that this correspondence notifies you of unauthorized use, and we reserve the right to pursue all remedies for willful infringement if the image continues to be used without a license.

If you have questions, please visit our FAQ [here].

On behalf of the 200,000 artists and photographers represented by [redacted], we thank you for your cooperation and look forward to assisting you however we can.

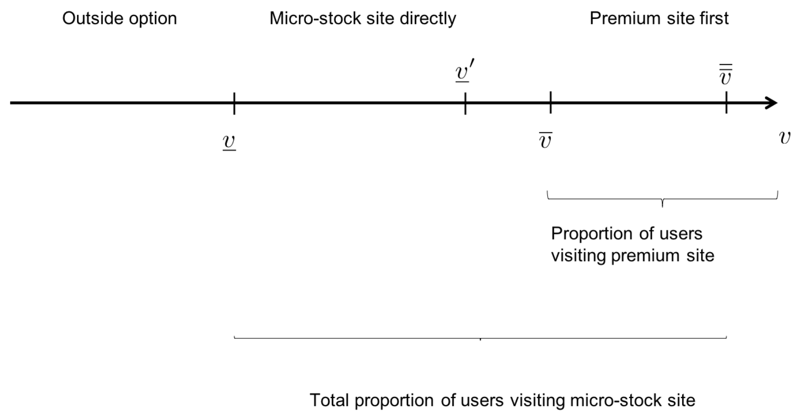
Sincerely,

Copyright Compliance Team
[redacted]
Reference #[Usage ID]

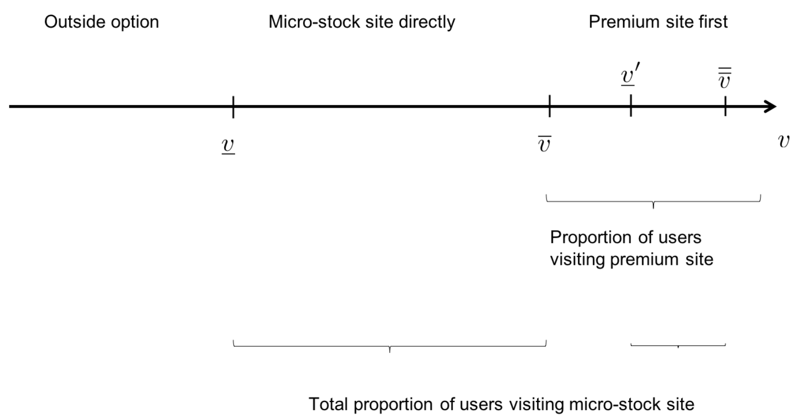
Notes: treatment group "Micro + Rec," with image recommendation but without premium-image price information. Under option 3, the email displays both a thumbnail and a link to the licensing page for each of the recommended images.

Figure 2: Illustration of the thresholds and the probabilities of searching either site

(a) Scenario 1: relatively small c_3 (the benchmark case is $c_3 = 0$)



(b) Scenario 2: sufficiently large c_3



Notes: These graphs illustrate the proportion of users visiting each of the two sites. $\underline{v}$ and $\overline{v}$ are the two thresholds separating the three initial search choices (Proposition 1): visit the premium site first; search the micro-stock site directly; and the outside option. $\overline{\overline{v}}$ is the threshold user who is indifferent between purchasing the premium image after discovering the actual premium price and the remaining options of continuing to search the micro-stock site or taking the outside option; and $\underline{v}'$ is the threshold user who is indifferent between continuing to search the micro-stock site after visiting the premium site and the outside option. In the first scenario (that is, $\overline{v} > \underline{v}'$), all users with $\overline{v} < v < \overline{\overline{v}}$ would find it worthwhile to search the micro-stock site after visiting the premium site. In the second scenario (that is, $\overline{v} < \underline{v}'$), some users do not find it worthwhile to visit the micro-stock site after spending time on the premium site. Corollary 1 provides a sufficient condition for Scenario 2.

Table 1: Experimental design

	Without premium-image price information	With premium-image price information
Premium option only	0	Price
Add micro-stock option	Micro	Micro + Price
Add micro-stock option + recommend images	Micro + Rec	Micro + Rec + Price

Table 2: Balance tests for cases in the analysis sample

	N	Multi-image case	High resolution	Secondary page	Case age	log(Total stock images on site + 1)
0	840	0.14	0.68	0.82	14.62	1.50
Price	911	0.12	0.67	0.79	14.46	1.49
		(0.31)	(0.76)	(0.12)	(0.43)	(0.82)
Micro	865	0.13	0.67	0.80	14.80	1.41
		(0.92)	(0.57)	(0.14)	(0.39)	(0.22)
Micro + Rec	1,947	0.11	0.68	0.80	14.50	1.51
		(0.02)	(0.91)	(0.17)	(0.49)	(0.93)
Micro + Price	897	0.10	0.68	0.80	14.48	1.53
		(0.03)	(0.87)	(0.14)	(0.48)	(0.68)
Micro + Rec + Price	1,947	0.12	0.68	0.81	14.44	1.56
		(0.13)	(0.96)	(0.50)	(0.32)	(0.36)
Total	7,407	0.12	0.68	0.80	14.53	1.51

Notes: the analysis sample. p-values in parentheses are based on two-sided t-tests between a given group and control group "0." Multi-image case indicates whether the case involves the unauthorized use of multiple images represented by the Agency; high resolution indicates whether the displayed image is high-resolution; secondary page equals one if the image is displayed on a secondary page of the firm's website rather than on the home page; case age is the number of months between the date when the case was identified and the date when the email was sent; and the number of total stock images on a website captures the number of unique images that are identified by a service provider as stock images represented by its client stock-photo agencies.

Table 3: Average effects of awareness of the micro-stock option

	N	License	Revenue	If license N	If license Revenue
		(1)	(2)		(3)
Two control groups	1,751	0.002	0.384	3	224.333
Four treatment groups	5,656	0.026	0.920	149	34.906
(p-value)		(0.000)	(0.218)		(0.002)

Notes: 14-day licensing outcomes of the two control groups and the four treatment groups in the analysis sample. We pool the groups together to estimate the average effects of increasing the awareness of the micro-stock licensing option. License equals one if the user purchases either a premium or a micro-stock license, and revenue (in \$) is the licensed revenue (\$0 if there is no license). p-values in parentheses are based on two-sided t-tests.

Table 4: Effects of the two interventions on micro-stock licensing: regression results

	License	Revenue	License	Revenue
	(1)	(2)	(3)	(4)
With image recommendations	0.008*	0.217**	0.010*	0.248**
	(0.004)	(0.102)	(0.005)	(0.113)
With premium-image price info	0.006	0.181*	0.009	0.224
	(0.004)	(0.106)	(0.007)	(0.146)
With image recommendations × With premium-image price info			-0.004	-0.061
			(0.008)	(0.201)
log(total stock images+1)	0.005***	0.115***	0.005***	0.115***
	(0.002)	(0.041)	(0.002)	(0.041)
Case age	-0.001	-0.015	-0.001	-0.015
	(0.000)	(0.012)	(0.000)	(0.012)
Multi-image case	0.006	0.164	0.006	0.166
	(0.008)	(0.218)	(0.008)	(0.217)
High resolution	-0.004	-0.093	-0.004	-0.094
	(0.005)	(0.112)	(0.005)	(0.112)
Secondary page	0.002	-0.120	0.002	-0.119
	(0.005)	(0.142)	(0.005)	(0.142)
Batch fixed effects	Y	Y	Y	Y
N	5656	5656	5656	5656

Notes: 14-day licensing outcomes of the affordable, micro-stock site for the four treatment groups in the analysis sample. License equals one if the user purchases a micro-stock license, and revenue (in \$) is the licensed revenue (\$0 if there is no license). "With image recommendations" equals one if emails belong to treatment groups "Micro + Rec" or "Micro + Price + Rec;" and "with premium-image price info" equals one if in groups "Micro + Price" or "Micro + Price + Rec." Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Effects of the two interventions on clicking onto the licensing sites: regression results

	Premium site	Micro-stock site	Premium site	Micro-stock site
	(1)	(2)	(3)	(4)
With image recommendations	0.017**	0.055***	0.022*	0.069***
	(0.008)	(0.007)	(0.013)	(0.010)
With premium-image price info	-0.088***	0.016**	-0.080***	0.036***
	(0.007)	(0.008)	(0.012)	(0.011)
With image recommendations			-0.011	-0.029**
×With premium-image price info			(0.015)	(0.015)
log(total stock images + 1)	0.009***	0.012***	0.009***	0.012***
	(0.002)	(0.003)	(0.002)	(0.003)
Case age	-0.002***	-0.001	-0.002***	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Multi-image case	-0.011	-0.020*	-0.011	-0.020*
	(0.011)	(0.012)	(0.011)	(0.012)
High resolution	0.024***	0.011	0.024***	0.011
	(0.008)	(0.008)	(0.008)	(0.008)
Secondary page	-0.025**	-0.019*	-0.025**	-0.018*
	(0.010)	(0.010)	(0.010)	(0.010)
Batch fixed effects	Y	Y	Y	Y
N	5656	5656	5656	5656

Notes: 14-day click-through rates onto the two licensing sites for the four treatment groups in the analysis sample. "With image recommendations" equals one if emails belong to treatment groups "Micro + Rec" or "Micro + Price + Rec;" and "with premium-image price info" equals one if in groups "Micro + Price" or "Micro + Price + Rec." Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendices (not for publication)

A. Appendix tables and figures

Table A1: Number of cases sent by group and by batch

	N	Batch 1 11/15/2017	Batch 2 11/20/2017	Batch 3 11/29/2017	Batch 4 12/04/2017
0	2,843	725	723	700	695
Price	2,848	726	725	700	697
Micro	2,839	721	726	692	700
Micro + Rec	6,359	1,627	1,625	1,552	1,555
Micro + Price	2,850	726	728	694	702
Micro + Rec + Price	6,351	1,624	1,622	1,553	1,552
Total	24,090	6,149	6,149	5,891	5,901

Notes: Includes all cases sent. As explained in detail in section B.3, we removed about four percent of the cases for the third and fourth batches because the Agency adjusted the tiering system of the firms during the experiment. As a result, some of the firms were no longer eligible for the experiment. The removed cases are statistically similar across different groups in their characteristics and proportional in quantity.

Table A2: Balance tests for cases sent

	N	Multi-image case	High resolution	Secondary page	Case age	log(Total stock images on site + 1)
0	2,843	0.13	0.67	0.75	14.93	1.21
Price	2,848	0.13	0.65	0.75	14.87	1.22
		(0.89)	(0.10)	(0.49)	(0.60)	(0.83)
Micro	2,839	0.13	0.66	0.73	14.96	1.18
		(0.65)	(0.44)	(0.11)	(0.80)	(0.54)
Micro + Rec	6,359	0.11	0.68	0.76	14.84	1.24
		(0.13)	(0.17)	(0.40)	(0.37)	(0.26)
Micro + Price	2,850	0.12	0.65	0.74	15.09	1.23
		(0.45)	(0.33)	(0.22)	(0.17)	(0.60)
Micro + Rec + Price	6,351	0.12	0.67	0.75	14.88	1.22
		(0.18)	(0.71)	(0.86)	(0.60)	(0.66)
Total	24,090	0.12	0.67	0.75	14.91	1.22

Notes: Includes all cases sent. p-values in parentheses are based on two-sided t-tests between a given group and group "0" in the first row. Multi-image case indicates whether the case involves the unauthorized use of multiple images represented by the Agency; high resolution indicates whether the displayed image is high-resolution; secondary page equals one if the image is displayed on a secondary page of the firm's website rather than on the home page; case age is the number of months between the date when the case was identified and the date when the email was sent; and the number of total stock images on a website captures the number of unique images identified by a service provider as stock images represented by its client stock-photo agencies.

Table A3: Generating the analysis sample

Group	N (sent)	Bounced back	N (not bounced)	Email opened in 14 days Percentage	Email opened in 14 days N
0	2,843	0.13	2,462	0.34	840
Price	2,848	0.13	2,488	0.37	911
		(0.39)		(0.07)	
Micro	2,839	0.13	2,468	0.35	865
		(0.71)		(0.49)	
Micro + Rec	6,359	0.13	5,529	0.35	1,947
		(0.65)		(0.34)	
Micro + Price	2,850	0.13	2,475	0.36	897
		(0.79)		(0.12)	
Micro + Rec + Price	6,351	0.13	5,548	0.35	1,947
		(0.32)		(0.40)	
Total	24,090	0.13	20,970	0.35	7,407

Notes: We removed cases from our analysis for which (1) the email was bounced back due to an invalid address; and (2) the email was not opened in the first 14 days. p-values in parentheses are based on two-sided t-tests between a given group and control group "0" in the first row.

B. Further details of the experiment

B.1 Email templates

Figure B1. Template illustration—emails to group "0"

SUBJECT LINE: Unauthorized image use on your site
PREHEADER: Action required for resolution

EMAIL:
Dear [COMPANY NAME],

We are contacting you because the below imagery represented exclusively by [redacted], a global digital media provider, is being displayed on your company's website [BUSINESS URL]:
[PREMIUM IMAGE THUMBNAIL IMAGE] [PREMIUM IMAGE THUMBNAIL SCREEN SHOT]

We are thrilled you like our imagery! However, we have no record of an active license for its commercial use on your site. We understand the unlicensed use may be accidental, but in fairness to our photographers, we ask that you take one of the following actions:

1. **Provide proof of a license that covers your use**
   If you do have a license, or if you believe you have mistakenly received this email, please email your licensing or other relevant information to [redacted], and indicate the reference number [Usage ID] in your e-mail.

2. **Purchase a license here** [PREMIUM IMAGE LICENSING PAGE] **to cover future use**
   This image is only available under a premium Rights Managed license.

In order for us to associate your purchase with this incident, please use the email address to which this notice was sent to create an account, and enter the reference number [Usage ID] in the "Purchase order number" field (click "Add note") when placing your order.

Please understand that this correspondence notifies you of unauthorized use, and we reserve the right to pursue all remedies for willful infringement if the image continues to be used without a license.

If you have questions, please visit our FAQ [here].

On behalf of the 200,000 artists and photographers represented by [redacted], we thank you for your cooperation and look forward to assisting you however we can.

Sincerely,

Copyright Compliance Team
[redacted]
Reference #[Usage ID]

Notes: control group, without premium-image price information.

Figure B2. Template illustration—emails to group “Micro”

SUBJECT LINE: Unauthorized image use on your site
PREHEADER: Action required for resolution

EMAIL:
Dear *[COMPANY NAME]*,

We are contacting you because the below imagery represented exclusively by [redacted], a global digital media provider, is being displayed on your company's website [BUSINESS URL]:
[RM THUMBNAIL IMAGE] [RM THUMBNAIL SCREEN SHOT]

We are thrilled you like our imagery! However, we have no record of an active license for its commercial use on your site. We understand the unlicensed use may be accidental, but in fairness to our photographers, we ask that you take one of the following actions:

1. **Provide proof of a license that covers your use**
   If you do have a license, or if you believe you have mistakenly received this email, please email your licensing or other relevant information to [redacted], and indicate the reference number [Usage ID] in your e-mail.

2. **Purchase a license here** [RM IMAGE ADP PAGE] **to cover future use**
   This image is only available under a premium Rights Managed license.

3. **Replace the premium image with a high-quality, but more affordable, image from** [redacted]
   Our [redacted] site offers a large collection of images that can be used for commercial websites with costs as low as $12 per image.

In order for us to associate your purchase with this incident, please use the email address to which this notice was sent to create an account, and enter the reference number [Usage ID] in the “Purchase order number” field (click “Add note”) when placing your order.

Please understand that this correspondence notifies you of unauthorized use, and we reserve the right to pursue all remedies for willful infringement if the image continues to be used without a license.

If you have questions, please visit our FAQ [here].

On behalf of the 200,000 artists and photographers represented by [redacted], we thank you for your cooperation and look forward to assisting you however we can.

Sincerely,

Copyright Compliance Team
[redacted]
Reference #[Usage ID]

Notes: treatment group, without image recommendation or premium-image price information.

Figure B3. Template illustration—emails to group "Price"

SUBJECT LINE: Unauthorized image use on your site
PREHEADER: Action required for resolution

EMAIL:
Dear *[COMPANY NAME]*,

We are contacting you because the below imagery represented exclusively by [redacted], a global digital media provider, is being displayed on your company's website [BUSINESS URL]:
[RM THUMBNAIL IMAGE] [RM THUMBNAIL SCREEN SHOT]

We are thrilled you like our imagery! However, we have no record of an active license for its commercial use on your site. We understand the unlicensed use may be accidental, but in fairness to our photographers, we ask that you take one of the following actions:

1. **Provide proof of a license that covers your use**
   If you do have a license, or if you believe you have mistakenly received this email, please email your licensing or other relevant information to [redacted], and indicate the reference number [Usage ID] in your e-mail.

2. **Purchase a license here** [PREMIUM IMAGE LICENSING PAGE] **to cover future use**
   This image is only available under a premium Rights Managed license. Licensing costs for online use of Rights Managed images typically range from $545 to $1140 for a 3-month period.

In order for us to associate your purchase with this incident, please use the email address to which this notice was sent to create an account, and enter the reference number [Usage ID] in the "Purchase order number" field (click "Add note") when placing your order.

Please understand that this correspondence notifies you of unauthorized use, and we reserve the right to pursue all remedies for willful infringement if the image continues to be used without a license.

If you have questions, please visit our FAQ [here].

On behalf of the 200,000 artists and photographers represented by [redacted], we thank you for your cooperation and look forward to assisting you however we can.

Sincerely,

Copyright Compliance Team
[redacted]
Reference #[Usage ID]

Notes: control group, with premium-image price information.

Figure B4. Template illustration—emails to group "Micro + Price"

SUBJECT LINE: Unauthorized image use on your site
PREHEADER: Action required for resolution

EMAIL:
Dear *[COMPANY NAME]*,

We are contacting you because the below imagery represented exclusively by [redacted], a global digital media provider, is being displayed on your company's website [BUSINESS URL]:
[RM THUMBNAIL IMAGE] [RM THUMBNAIL SCREEN SHOT]

We are thrilled you like our imagery! However, we have no record of an active license for its commercial use on your site. We understand the unlicensed use may be accidental, but in fairness to our photographers, we ask that you take one of the following actions:

1. **Provide proof of a license that covers your use**
   If you do have a license, or if you believe you have mistakenly received this email, please email your licensing or other relevant information to [redacted], and indicate the reference number [Usage ID] in your e-mail.

2. **Purchase a license here** [RM IMAGE ADP PAGE] **to cover future use**
   This image is only available under a premium Rights Managed license. Licensing costs for online use of Rights Managed images typically range from $545 to $1140 for a 3-month period.

3. **Replace the premium image with a high-quality, but more affordable, image from** [redacted]
   Our [redacted] site offers a large collection of images that can be used for commercial websites with costs as low as $12 per image.

In order for us to associate your purchase with this incident, please use the email address to which this notice was sent to create an account, and enter the reference number [Usage ID] in the "Purchase order number" field (click "Add note") when placing your order.

Please understand that this correspondence notifies you of unauthorized use, and we reserve the right to pursue all remedies for willful infringement if the image continues to be used without a license.

If you have questions, please visit our FAQ [here].

On behalf of the 200,000 artists and photographers represented by [redacted], we thank you for your cooperation and look forward to assisting you however we can.

Sincerely,

Copyright Compliance Team
[redacted]
Reference #[Usage ID]

Notes: treatment group, without image recommendation but with premium-image price information.

Figure B5. Template illustration—emails to group "Micro + Rec + Price"

SUBJECT LINE: Unauthorized image use on your site
PREHEADER: Action required for resolution

EMAIL:
Dear *[COMPANY NAME]*,

We are contacting you because the below imagery represented exclusively by [redacted], a global digital media provider, is being displayed on your company's website [BUSINESS URL]:
[RM THUMBNAIL IMAGE] [RM THUMBNAIL SCREEN SHOT]

We are thrilled you like our imagery! However, we have no record of an active license for its commercial use on your site. We understand the unlicensed use may be accidental, but in fairness to our photographers, we ask that you take one of the following actions:

1. **Provide proof of a license that covers your use**
   If you do have a license, or if you believe you have mistakenly received this email, please email your licensing or other relevant information to [redacted], and indicate the reference number [Usage ID] in your e-mail.

2. **Purchase a license here** [RM IMAGE ADP PAGE] **to cover future use**
   This image is only available under a premium Rights Managed license. Licensing costs for online use of Rights Managed images typically range from $545 to $1140 for a 3-month period.

3. **Replace the premium image with a high-quality, but more affordable, image from** [redacted]
   Our [redacted] site offers a large collection of images that can be used for commercial websites with costs as low as $12 per image. We have selected the images below for you as possible replacements:
   [IMAGE URL SIMILAR 1] [IMAGE URL SIMILAR 2] [IMAGE URL SIMILAR 3] [IMAGE URL SIMILAR 4]

In order for us to associate your purchase with this incident, please use the email address to which this notice was sent to create an account, and enter the reference number [Usage ID] in the "Purchase order number" field (click "Add note") when placing your order.

Please understand that this correspondence notifies you of unauthorized use, and we reserve the right to pursue all remedies for willful infringement if the image continues to be used without a license.

If you have questions, please visit our FAQ [here].

On behalf of the 200,000 artists and photographers represented by [redacted] we thank you for your cooperation and look forward to assisting you however we can.

Sincerely,

Copyright Compliance Team
[redacted]
Reference #[Usage ID]

Notes: treatment group, with both image recommendation and premium-image price information. Under option 3, the email displays both a thumbnail and a link to the licensing page for each of the recommended images.

B.2 Case eligibility

At the Agency, infringing firms are categorized into five 'tiers,' using a formula developed by a third-party data firm. The formula uses information about each infringing firm—such as annual revenues, number of employees, and industry—in order to predict the likelihood that the firm will settle an infringement claim at the list price of a premium image. Tier 1 is the most likely to settle and tier 5 the least likely. We do not observe the formula that maps a firm's observable characteristics to a particular tier, but firms with lower annual revenues (or if the revenue information is missing) tend to have a higher tier number. In a different experiment, Luo and Mortimer (2017) show that the likelihood of settlement increases significantly with a firm's annual revenue (or, equivalently, decreases significantly with the tier number).

As of January 2014, the Agency no longer requested any settlement amount from firms in tiers 4 and 5 but continued to pursue settlement from firms in lower tiers (larger firms). The new experiment that this paper analyzes included only firms in tiers 4 and 5 (that is, the smallest firms) to avoid disrupting the normal operation of the compliance team at the Agency. The cases were identified within the two years preceding the experiment.

B.3 Allocation

The cases were allocated into the six groups in two steps. For about ten percent of the cases eligible for the experiment, the similar-image algorithm initially did not yield any results. To avoid confounding the effect of our interventions with the possibility that the user might not find similar replacement images on the website through their own searches, we excluded these cases at first. All other cases were randomly allocated using a random-number generator to the six groups, as presented in table 1. Specifically, we allocated one tenth of the cases to each of the four groups (two control and two treatment groups) that do not receive image recommendations, and three tenths to each of the two treatment groups that received recommendations. We intentionally allocated more cases to the two treatment groups for which similar images are recommended, as the Agency deemed this to be the most constructive approach. It turned out later that the similar-image algorithm had not initially worked due to a technical glitch, which was not systematically correlated with the likely number of similar images on the licensing site or the characteristics of the case. In order to preserve as many observations as possible and to provide a more balanced number of observations across groups, we reinstated these cases and randomly added them equally to the four groups that were not given image recommendations.[41]

We planned to send emails for 24,680 cases in four batches. The number of cases ranged from 726 to 728 per batch for the two control and two treatment groups without image recommendations, and from 1,630 to 1,631 per batch for the two treatment groups with image recommendations. As illustrated in table A1, the total number of emails that the Agency sent out was 24,090; thus, there is a discrepancy of 590 cases. 575 of these cases (98 percent) had been removed from the third and the fourth batches because the Agency changed its classification system for infringing firms in the middle of the experiment, which made about four percent of the cases ineligible for the experiment. We confirmed that the removed cases were statistically similar across different groups in their characteristics and were proportional in quantity. We were not able to trace the cause for the remaining 15 missing cases, and they appear to have come from different groups or batches without a systematic pattern. Table A2 shows that for cases that were eventually sent out, the groups were well-balanced.

B.4 Follow-up emails

A follow-up email was sent two weeks after the initial email, as long as there was no record of licensing events or email correspondence. The follow-up email was exactly the same as the initial email (that is, the interventions were consistent), except for the opening sentence, which indicated that it was a follow-up to the previous correspondence. There were no other consequences for continued infringement within the scope of this experiment.

[41] It is important to note that the glitch was discovered before the emails were sent. Thus, these recipients received exactly the email template that they were supposed to according to the group assignment. In other words, it is *not* the case that these firms received the 'Rec' templates and that we grouped them into the other four groups only in the analysis.

C. Search and potential mechanisms: a model

C.1 Proofs of Proposition 1 and Predictions 1 and 2

In this section, we provide proofs of the results from the model presented in Section 3.3 of the paper. Recall that $\overline{v}$ is the user who is indifferent between going to the micro-stock site directly and first visiting the premium site (that is, the solution to $U^{\mathrm{M}}(v) = U^{\mathrm{P}}(v)$); $\underline{v}$ is indifferent between visiting the micro-stock site directly and taking the outside option (that is, the solution to $U^{\mathrm{M}}(v) = 0$); $\underline{v}'$ is indifferent between continuing to search the micro-stock site and taking the outside option after spending time on the premium site (that is, the solution of $U^{\mathrm{M}}(v) - c_3 = 0$; and $\overline{\overline{v}} = \tilde{p} + m(v) = \tilde{p} + \max\{U^{\mathrm{M}}(v) - c_3, 0\}$ is the threshold user who finds it worthwhile to purchase the premium image after discovering the actual premium price, $\tilde{p}$.

As explained in the paper, we make the following assumptions to exclude scenarios that are not relevant to our empirical context. Assumption 1 guarantees that each of the three choices can be optimal for some values of v:

Assumption 1. $c_1 > \int_{\underline{P}}^{\underline{v}} (\underline{v} - p) f(p) dp.$

Specifically, the direct cost of searching for the premium-price information is sufficiently high such that a user who is indifferent between visiting the micro-stock site directly and taking the outside option, $\underline{v}$, does not find it worthwhile to visit the premium site first.[42]

Assumption 2 ensures that the actual price is sufficiently high that only a small percentage of users who visit the premium site find it worthwhile to purchase the premium image. Notice that $\overline{v}$ cannot exceed a certain upper bound, because $U^{\mathrm{P}}(v)$ is a non-decreasing function of $m(v)$, and $m(v)$ is bounded from below at zero. Denote this upper bound as $\overline{v}^{\max}$, and assume the following:[43]

Assumption 2. $\tilde{p} > \overline{v}^{max}$.

With Assumption 2, $\overline{\overline{v}} = \tilde{p} + m(v) > \overline{v}^{\max} + m(v) > \overline{v}$. Thus, only a small percentage of users who visit the premium site will purchase the premium image, which is consistent with our data.

Recall that Proposition 1 states that the user follows a threshold rule for the three initial search options described in equation (1): there exist unique $\underline{v} < \overline{v}$ such that the user would take the outside option if $v < \underline{v}$; go to the micro-stock site directly if $\underline{v} < v < \overline{v}$; and go to the premium site first if $v > \overline{v}$.

[42]Notice that $m(\underline{v}) = \max\{-c_3, 0\} = 0$. Thus, $U^{\mathrm{P}}(\underline{v}) = \int_{\underline{P}}^{\underline{v}} (\underline{v} - p) f(p) dp - c_1$.

[43]$\overline{v}^{\max}$ is the solution of $\int_{12/v}^{1} (vq - 12) g(q) dq - c_2 = \int_{\underline{P}}^{v} (v - p) f(p) dp - c_1$. The existence and uniqueness of $\overline{v}^{\max}$ can be proved using similar arguments in Proposition 1.

Proof of Proposition 1. First, $U^{\mathrm{M}}(v)$ is an increasing function of v. In particular, $U^{\mathrm{M}}(v)$ can be written as $\int_{12/v}^{1}(vq-12)g(q)dq-c_2$, and $\frac{\partial U^{\mathrm{M}}(v)}{\partial v}=\int_{12/v}^{1}qg(q)dq>0$. $\lim_{v\to 0}U^{\mathrm{M}}(v)=-c_2$ and $\lim_{v\to\infty}U^{\mathrm{M}}(v)=\infty$. Thus, there exists a unique $\underline{v}$ such that $U^{\mathrm{M}}(\underline{v})=0$.

To show that $\overline{v}$ exists and is unique, we first derive $\frac{\partial U^{\mathrm{P}}(v)}{\partial v}$. Notice that $U^{\mathrm{P}}(v)=m(v)+\int_{\underline{P}}^{v-m(v)}(v-p-m(v))f(p)dp-c_1$. Moreover, denote by $\underline{v}'$ the threshold value that equates $U^{\mathrm{M}}(v)-c_3=0$ (that is, the user who is indifferent between visiting the micro-stock site and the outside option after visiting the premium site).[44] Because $m(v)=U^{\mathrm{M}}(v)-c_3$ for $v>\underline{v}'$ and 0, otherwise, the derivative of $\frac{\partial U^{\mathrm{P}}(v)}{\partial v}$ takes different expressions for these two segments:

$$\frac{\partial U^{\mathrm{P}}(v)}{\partial v}=\begin{cases}\frac{\partial U^{\mathrm{M}}(v)}{\partial v}+\int_{\underline{P}}^{v-U^{\mathrm{M}}(v)+c_3}(1-\frac{\partial U^{\mathrm{M}}(v)}{\partial v})f(p)dp & \text{if } v>\underline{v}'\\ F(v) & \text{if } v\leq\underline{v}'\end{cases}.$$

For each of the above two segments, $U^{\mathrm{P}}(v)$ is an increasing function of v. When $v\leq\underline{v}'$, the derivative is positive because $F(v)>0$. When $v>\underline{v}'$, $\frac{\partial U^{\mathrm{M}}(v)}{\partial v}+\int_{\underline{P}}^{v-U^{\mathrm{M}}(v)+c_3}(1-\frac{\partial U^{\mathrm{M}}(v)}{\partial v})f(p)dp=\frac{\partial U^{\mathrm{M}}(v)}{\partial v}+(1-\frac{\partial U^{\mathrm{M}}(v)}{\partial v})F(v-U^{\mathrm{M}}(v)+c_3)>0$, as $\frac{\partial U^{\mathrm{M}}(v)}{\partial v}=\int_{12/v}^{1}qg(q)dq<E_q[q]<1$.

For the existence of $\overline{v}$, notice two things. First, under Assumption 1, $U^{\mathrm{P}}(v)<U^{\mathrm{M}}(v)$ when $v=\underline{v}$. Second, $U^{\mathrm{P}}(v)$ increases at an increasing rate for $\underline{v}<v<\underline{v}'$ (because the derivative $F(v)$ is an increasing function of v); when $v=\underline{v}'$, $\frac{\partial U^{\mathrm{P}}(v)}{\partial v}$ increases from $F(\underline{v}')$ to $\frac{\partial U^{\mathrm{M}}(\underline{v}')}{\partial v}+(1-\frac{\partial U^{\mathrm{M}}(\underline{v}')}{\partial v})F(\underline{v}')$; and $\frac{\partial U^{\mathrm{P}}(v)}{\partial v}>\frac{\partial U^{\mathrm{M}}(v)}{\partial v}$ for any $v>\underline{v}'$ (as shown in the previous paragraph). Thus, there must exist a value of $v>\underline{v}$ such that $U^{\mathrm{P}}(v)$ crosses $U^{\mathrm{M}}(v)$ from below.

For the uniqueness of $\overline{v}$, the goal is to show that once $U^{\mathrm{M}}(v)$ and $U^{\mathrm{P}}(v)$ cross at $\overline{v}$, they will not cross again. This is guaranteed if their difference, $\Delta(v)=U^{\mathrm{P}}(v)-U^{\mathrm{M}}(v)$, is monotonically increasing in v after $\overline{v}$. Take the derivative of the difference w.r.t. v:

$$\frac{\partial\Delta(v)}{\partial v}=\begin{cases}\int_{\underline{P}}^{v-U^{\mathrm{M}}(v)+c_3}(1-\frac{\partial U^{\mathrm{M}}(v)}{\partial v})f(p)dp & \text{if } v>\underline{v}'\\ F(v)-\frac{\partial U^{\mathrm{M}}(v)}{\partial v} & \text{if } v\leq\underline{v}'\end{cases}.$$

If $\overline{v}\geq\underline{v}'$, because the first line of the above expression is always positive, $\Delta(v)$ increases with v for $v>\overline{v}$. Thus, $U^{\mathrm{M}}(v)$ and $U^{\mathrm{P}}(v)$ will not cross again.

If $\overline{v}<\underline{v}'$, the following shows that $\frac{\partial\Delta(v)}{\partial v}>0$ for any $\overline{v}<v\leq\underline{v}'$. Combined with the result in the previous

[44] Similar to $\underline{v}$, $\underline{v}'$ exists and is unique. Furthermore, $\underline{v}'\geq\underline{v}$.

paragraph for $v > \overline{v}$, we also have $U^{\mathrm{M}}(v)$ and $U^{\mathrm{P}}(v)$ never crossing again. To show that $\frac{\partial \Delta(v)}{\partial v} > 0$ for any $\overline{v} < v \leq \underline{v}'$, consider an alternative payoff function, $\hat{U}^{\mathrm{M}}(v) = E_q[q] * v - 12 - c_2$; that is, the user always pays \$12 (and the search cost c_2) and receives an image value of the mean expected quality. Notice that $\hat{U}^{\mathrm{M}}(v)$ is everywhere below $U^{\mathrm{M}}(v)$ because, with the latter, the downside risk that the quality yields the user a value lower than \$12 is eliminated by the option of not buying the image. Thus, $\hat{U}^{\mathrm{M}}(v)$ must cross $U^{\mathrm{P}}(v)$ at a value of $\hat{\overline{v}} < \overline{v}$. Note that the slope of $\hat{U}^{\mathrm{M}}(v)$, which is $E_q[q]$, is constant and that the slope of $U^{\mathrm{P}}(v)$, which is $F(v)$, increases with v. Thus, the difference in the slopes, $F(v) - E_q[q]$, is positive after $\hat{\overline{v}}$. Finally, recall that $\frac{\partial U^{\mathrm{M}}(v)}{\partial v} < E_q[q]$ as shown above. Thus, $\frac{\partial \Delta(v)}{\partial v} = F(v) - \frac{\partial U^{\mathrm{M}}(v)}{\partial v} > F(v) - E_q[q] > 0$ for $\overline{v} < v \leq \underline{v}'$. □

Equation (3) in the paper presents the probability of users searching the micro-stock sites, and it considers two scenarios. The following corollary provides a sufficient condition for the existence of the second scenario; that is, $\overline{v} < \underline{v}'$.

Corollary 1. *There exists a unique $\bar{c}_3$ such that when $c_3 > \bar{c}_3$, among users who visit the premium site first, at least some will find it too costly to visit the micro-stock site afterwards.*

Proof. As explained above, $\overline{v}$ can not exceed $\overline{v}^{\max}$. At the same time, by the implicit function theorem and the fact that $U^{\mathrm{M}}(v) - c_3$ increases with v and decreases with c_3, $\underline{v}'$ (the solution to $U^{\mathrm{M}}(v) = c_3$) is an increasing function of c_3 and is unbounded from above. Thus, a sufficient condition for $\overline{v} < \underline{v}'$ is when $c_3 > \bar{c}_3$, where $\bar{c}_3 = U^{\mathrm{M}}(\overline{v}^{\max})$. □

In the following, we provide the proofs of the two predictions on the effects of the two interventions—(i) image recommendations; and (ii) provision of the premium-image price information—on the probability of searching the premium site, S^{P}, and the probability of searching the micro-stock site, S^{M}.

Proof of Prediction 1.

Proof for (i.a):

When $c_3 = 0$, $\underline{v}' = \underline{v} < \overline{v}$. Thus, $\mathrm{S}^{\mathrm{M}} = \Pr(\underline{v} \leq v \leq \overline{\overline{v}})$ (see equation (3)). $\underline{v}$, the solution of $U^{\mathrm{M}}(v) = 0$, is an increasing function of c_2 by the implicit function theorem, as the left-hand side of this equation is an increasing function of v and a decreasing function of c_2. $\overline{\overline{v}} = \tilde{p} + \max\{U^{\mathrm{M}}(v), 0\}$ is a non-increasing function of c_2. Thus, as c_2 decreases, $\underline{v}$ decreases and $\overline{\overline{v}}$ weakly increases. As a result, S^{M} increases.

$S^P = \Pr(v > \overline{v})$ (see equation (2)). When $c_3 = 0$, $\overline{v}$ is determined by the following equation:[45]

$$\Delta(v) = U^P(v) - U^M(v) = \int_{\underline{P}}^{v-U^M(v)} (v - p - U^M(v))f(p)dp - c_1 = 0.$$

$\frac{\partial \Delta(v)}{\partial v} = \int_{\underline{P}}^{v-U^M(v)} (1 - \frac{\partial U^M(v)}{\partial v})f(p)dp > 0$ and $\frac{\partial \Delta(v)}{\partial c_2} = \int_{\underline{P}}^{v-U^M(v)} -\frac{\partial U^M(v)}{\partial c_2} f(p)dp > 0$. By the implicit function theorem, $\overline{v}$ is a decreasing function of c_2. Thus, when c_2 decreases, $\overline{v}$ will increase. As a result, S^P decreases.

Proof for (i.b):

When $c_3 > 0$, $\overline{v}$ weakly decreases as c_3 decreases. This is because $U^M(v)$ is not a function of c_3, while $U^P(v)$ weakly decreases with c_3 (as a greater c_3 implies a worse $m(v)$). Thus, image recommendations, via reducing c_3, also have a negative effect on $\overline{v}$, and, hence, a positive effect on S^P. The total effect of image recommendations on S^P, therefore, is the sum of the negative effect from a reduction in c_2 (as shown above in (i.a)) and the positive effect from a reduction in c_3. If the latter, positive effect is sufficiently large (e.g., if c_3 is very high without the intervention and drops to zero with this intervention) to overcome the former cannibalization effect, we will have a lower $\overline{v}$ and a higher S^P.

Proof for (ii):

The revealed image quality either increases or decreases the expected *benefit* of searching for a replacement image (that is, the $\Pr(vq \geq 12)E_q[vq - 12 | vq \geq 12]$ part of $U^M(v)$) and, hence, $U^M(v)$ itself. Suppose that the new quality information yields a higher $U^M(v)$, it is straightforward that this will lead to a lower $\underline{v}$ and a lower $\underline{v}'$.

Recall that $\overline{v}$ is the solution of the following equation:

$$\Delta(v) = m(v) - U^M(v) + \int_{\underline{P}}^{v-m(v)} (v - p - m(v))f(p)dp - c_1 = 0.$$

As shown in the proof of Proposition 1, $\frac{\partial \Delta(v)}{\partial v} > 0$ at $v = \overline{v}$. To see the sign of $\frac{\partial \Delta(v)}{\partial U^M(v)}$, consider two different scenarios. First, when $m(v) = \max\{U^M(v) - c_3, 0\} = 0$, $\Delta(v) = -U^M(v) + \int_{\underline{P}}^{v}(v - p)f(p)dp - c_1$. We can show that $\Delta(v)$ decreases with $U^M(v)$. Second, if $m(v) = \max\{U^M(v) - c_3, 0\} = U^M(v) - c_3$, we have $\Delta(v) = -c_3 + \int_{\underline{P}}^{v-U^M(v)+c_3} (v - p - U^M(v) + c_3))f(p)dp - c_1$, which we can also show to be a decreasing function of $U^M(v)$. Thus, $\overline{v}$ increases as $U^M(v)$ increases due to the implicit function theorem.

Finally, note that $\overline{\overline{v}} = \tilde{p} + m(v) = \tilde{p} + \max\{U^M(v) - c_3, 0\}$. Thus, as $U^M(v)$ increases, $\overline{\overline{v}}$ either increases

[45] Under Assumption 1, $m(v) = \max\{U^M(v), 0\} = U^M(v)$ for $v > \underline{v}$, which is the relevant range of v's to consider here.

or does not change.

Prediction ii.a shows that if the revealed quality of the micro-stock image is higher than expected, we should expect to see a higher S^M and a lower S^P. This prediction does not depend on the value of c_3. To see this, consider the two scenarios illustrated in Figure 2.

Consider the first scenario as illustrated in Figure 2a (when c_3 is relatively small). The probability of searching the micro-stock site is $S^M = \Pr(\underline{v} \le v \le \overline{\overline{v}})$. If the revealed quality of the micro-stock image is higher than expected, S^M increases, because $\underline{v}$ decreases while $\overline{\overline{v}}$ either increases or does not change. The probability of searching the premium site is $S^P = \Pr(v > \overline{v})$, which will decrease because $\overline{v}$ increases.

Consider the second scenario as illustrated in Figure 2b (when c_3 is sufficiently high). The probability of searching the micro-stock site is then $\Pr(\underline{v} \le v < \overline{v}) + \Pr(\underline{v}' \le v \le \overline{\overline{v}})$. Both segments of this total probability increase because $\underline{v}$ and $\underline{v}'$ both become lower, $\overline{v}$ becomes higher, and $\overline{\overline{v}}$ becomes higher or does not change. $S^P = \Pr(v > \overline{v})$ decreases as explained above.

Prediction ii.b can be shown similarly. □

Proof of Prediction 2.

Proof for (i):

With $\tilde{p}$ revealed, the user compares the payoff from purchasing a premium license and the maximum of the other two options (finding a replacement image and the outside option). Thus, the proportion of users going to the premium site will be $S^P = \Pr(v \ge \tilde{p} + \max\{U^M(v), 0\})$. Note that because the user will not have incurred c_1 before knowing the price information, if she is to search the micro-stock site, she needs to incur only c_2 instead of $c_2 + c_3$. Because $\tilde{p} + \max\{U^M(v), 0\} \ge \tilde{p} + \max\{U^M(v) - c_3, 0\} = \overline{\overline{v}}$ and that $\overline{\overline{v}} > \overline{v}$ (Assumption 2), the probability of visiting the premium site with price revealed, $\Pr(v \ge \tilde{p} + \max\{U^M(v), 0\})$, is lower than that without the price revealed, which is $\Pr(v \ge \overline{v})$.

Proof for (ii):

If c_3 is 0 to start with, $S^M = \Pr(\underline{v} \le v \le \overline{\overline{v}})$ (equation (3)). Revelation of the premium-price information in the email does not affect either threshold, thus, S^M should stay the same. When $c_3 > 0$, however, immediate revelation of the premium price will guide users who cannot afford that price to go to the micro-stock site directly, without wasting time searching for the premium price; that is, users will face c_2 rather than $c_2 + c_3$ in searching for a replacement image. Thus, $S^M = \Pr(\underline{v} < v < \tilde{p} + \max\{U^M(v), 0\})$. $\underline{v}$ does not change with this intervention, but $\tilde{p} + \max\{U^M(v), 0\}$ is greater than or equal to the old $\overline{\overline{v}}$, which is $\tilde{p} + \max\{U^M(v) - c_3, 0\}$

in the absence of this intervention. Thus, S^{M} weakly increases. In particular, if c_3 is sufficiently high for us to be in the scenario described in Figure 2b, the increase in S^{M} will be strict because the gap between $\overline{v}$ and $\underline{v}'$ also disappears. $\square$

D. Longer-term outcomes

To have a sense about users' non-licensing behaviors that we group together as the outside option in the paper (e.g., continual infringement or taking down the images with or without replacement), we collected longer-term outcomes in May 2019, about a year and a half after the start of our experiment.

We randomly selected 100 cases from each of the six groups (a total of 600 cases, about eight percent of our analysis sample) and manually collected the following variables: (i) whether the infringed image is still displayed on the page URL on which the infringement was first detected; and (ii) whether we can find any image on the same page that looks similar (by visual examination) to the infringed premium image but is not licensed through the Agency. The second variable is meant to capture incidences in which the users replace with a similar image from outside sources. However, this is, at best, a noisy measure for the following reasons. First, without knowing when the image files were coded into the webpages, we cannot tell whether they were already on display before our email interventions. Second, even if these images were placed on the webpages after our emails, we cannot tell whether they were used as a response to our emails or for unrelated reasons. Third, even though the first two reasons may over-estimate the replacement probability as a response to our emails, there are also reasons to believe that we may not capture the full extent of such behaviors—for example, if firms do not have to replace the infringed images with ones that look similar.

Another issue that we encounter during the data-collection process is that about 40 percent of pages are no longer valid.[46] Similar to our discussion above, some of these pages might already have been invalid before our emails, which is possible given that the average age of the case (from detection to our email) is 14.5 months. Those that were made invalid afterwards might have happened in response to our emails or for unrelated reasons. Panel (c) in table A4 presents the probability of invalid pages by group; pair-wise test statistics show that this probability does not differ across groups.[47]

In panels (a) and (b) of table A4, we present the probability of continual infringement and the probability of our measure of 'image replacement (from outside sources)' for all six groups, for all 600 cases. Overall, the rate of continual infringement is 12.5 percent; the rate of image replacement is 19.5 percent; and 28 percent took down the premium image without replacing it.[48]

[46]Recall that the majority of the pages are not the home page of a firm, so invalid pages do not mean that the firms are no longer in operation.

[47]Because our experiment uses randomization (and the probability of invalid pages is similar across groups), we do not think that these invalid pages affect the interpretation of our licensing results in the paper. If anything, the possibility of invalid pages prior to our email interventions suggests that the licensing rate is likely to be higher if emails were sent out in a more timely manner.

[48]These percentages are not conditional on a page remaining active. For the conditional statistics, one should divide these rates

The differences across groups are not statistically significant, with the exception of the comparison between the two treatment groups that receive image recommendations but do or do not receive the premium-price information. However, it is not clear why this might be the case. The data provide a noisy measure of the true underlying behavior, and the difference due to premium-price information in the other two comparisons is either zero (between the two treatment groups that do not receive image recommendations) or very small and insignificant (between the two control groups, p-value of 0.49).

Even though it is impossible to reach precise conclusions, these data do make us think further about the different outside options that users may choose. With all the data caveats in mind, here are some of our take-aways:

First, the continual infringement rate (12.5 percent) is not trivial. These are likely to be cases in which the users perceive a low legal risk of continual infringement (which is not surprising given the generally non-threatening tone of the emails); and they also receive some value from keeping the particular image, as it might be an integral part of the website design, whereas the costs of finding a replacement image and integrating it into the site are relatively high. Thus, for these users, taking no action (i.e., continual infringement) may be the least costly option.

Second, for the 28 percent of cases in which the image is taken down without being replaced, the users likely face non-trivial costs of finding and uploading a replacement image and/or do not place much value on using an (or an additional) image.

Lastly, the incidence of using a replacement image from other sources is 19.5 percent. This suggests that for a sizable proportion of users, the additional costs of finding and uploading replacement images from outside sources are not prohibitively high and that they derive sufficient value from having an additional image on their websites. It is possible that these users are sufficiently price-sensitive to take their own photographs, or they may use Creative Commons images (such as from Flickr.com) for free. That said, given that the Agency's portfolio of micro-stock images is among the largest in the industry and that they cost only about $20—which does not seem like much for even a small business—this raises an interesting question: why don't these users simply license from the Agency, as offered in our emails? It is possible that they might not have been able to find a replacement image on the Agency's site; or they may have reacted negatively to being monitored, particularly in the absence of an industry norm for ex-post licensing requests.

by about 0.6.

Table A4: Long-term outcomes

(a) Whether premium image still displayed on page

	Without premium-image price information	With premium-image price information
Premium option only	0.12	0.13
Add micro-stock option	0.16	0.13
Add micro-stock option + recommend images	0.09	0.12

(b) Whether a similar image is found on page that is not licensed from the Agency

	Without premium-image price information	With premium-image price information
Premium option only	0.24	0.20
Add micro-stock option	0.17	0.17
Add micro-stock option + recommend images	0.25	0.14

(c) Whether page URL is still valid

	Without premium-image price information	With premium-image price information
Premium option only	0.59	0.63
Add micro-stock option	0.62	0.63
Add micro-stock option + recommend images	0.58	0.60

Notes: a random subset of our analysis sample, with 100 cases from each group.