

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

Hong Luo<sup>y</sup>

Julie Holland Mortimer<sup>z</sup>

December 27, 2018

## Abstract

Copyright infringement may result from frictions preventing legal consumption, but may also reveal demand. Motivated by this fact, we run a field experiment in which 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 site. Making infringers aware of the cheaper option leads to a fourteen-fold increase in the ex-post licensing rate. Two additional experimental interventions are designed to reduce search costs for (i) price and (ii) product information. Both interventions—immediate price comparison and recommendation of images similar to those infringed—have large positive effects. Our results highlight the importance of mitigating user costs in small-value transactions.

---

The authors thank Tobias Kretschmer, Michael Luca, Shane Greenstein, Felix Oberholzer-Gee, and Martin Watzinger 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.

<sup>y</sup>Harvard Business School; email: hluo@hbs.edu.

<sup>z</sup>Boston College and NBER; julie.mortimer.2@bc.edu.

# 1 Introduction

Search engines and social media have made it easy for users to search for digital content or to encounter it serendipitously, and consumers' use of this content may be informative about their unobserved preferences for specific products. When legal options for purchasing content are difficult to access, or when licensing obligations are not sufficiently clear, consumers may possibly unwittingly violate copyright law. When this happens, infringement may provide demand discovery and, hence, serve as a marketing opportunity. Indeed, it may be less costly to monitor online infringement than to identify potential customers ex-ante through other means.

These observations suggest the possibility of approaching infringing users to (i) clarify their legal obligations; (ii) increase their awareness of more-suitable purchasing options; and (iii) mitigate their search and transactions costs (e.g., by recommending replacement products based on information revealed by their infringing use). Our paper explores the feasibility of an 'ex-post licensing' approach by private parties.<sup>1</sup> Naturally, from a copyright holder's perspective, the value of this approach depends on a number of factors, including the characteristics of the infringing users and the infringement, the costs of identifying infringement relative to that of predicting demand ex-ante, and the ability to motivate users to pay ex-post.<sup>2:3</sup> In a frictionless world, infringing users are typically thought of as price-sensitive users who would not purchase content legally. However, if frictions are sufficient to prevent legal consumption, it is possible that reducing these frictions may result in legal purchases.

Our empirical context is the stock-photography (pre-shot images) industry.<sup>4</sup> On behalf of photographers, stock-photo agencies license images to business customers. The industry is divided

---

<sup>1</sup>The idea that transactions costs are mitigated by allowing follow-on products to be created without obtaining permission from right holders is a key argument for establishing an alternative copyright regime (Lessig (2008); Gans (2015); Menell (2016)). Gans (2015) provides the first theoretical framework on the impacts of different remix/broadly defined as derivative works created by combining or editing existing materials/rights regimes on the incentives to create not just follow-on, but also original, content. Menell (2016) makes a legal case for a remix compulsory licensing regime.

<sup>2</sup>The approach we analyze is qualitatively different from strategies that aim to maximize enforcement revenues by exploiting a legal threat. See Scott Morton and Shapiro (2014) for enforcement strategies and the notion of reasonable royalties for patents.

<sup>3</sup>Digimarc and Thomson Reuters provide infringement-detection services for movies, music, software, and publishing. YouTube automates the detection of infringement and takes actions according to copyright owners' pre-specified choices, including sharing ad revenue from uploaded videos.

<sup>4</sup>According to one estimate, the global revenue of this industry was \$2.88B in 2011. In comparison, the royalties collected by the two largest public performance rights organizations in the music industry, ASCAP and BMI, were \$1B each in 2012.

into premium and micro-stock segments. The premium segment targets usage occasions such as

comparison increases the micro-stock licensing rate by 31 percent.

Data on whether users click onto either of the licensing sites suggest that both interventions are effective in reducing users' costs of finding a replacement image. Furthermore, the evidence suggests that image recommendation induces marginal users (i.e., those with a relatively low willingness to pay) to start a search instead of choosing the outside option; whereas providing premium price information may direct infra-marginal users (i.e., those with a relatively high willingness to pay) away from the unsuitable premium option faster.

Our results contribute to the literature on information provision and search costs, in particular in the context of electronic marketplaces. Both interventions appear to facilitate matching between users and their preferred products, consistent with studies on the effect of information provision on the quality of matching (Anand and Shachar (2011); Tadelis and Zettelmeyer (2015)) and the effects of reducing search costs on peer-to-peer platforms. See Fradkin (2017) and Horton (forthcoming) on providing better-updated information about the availability of homes on Airbnb or of workers on oDesk.

Our paper also relates to the sampling effect investigated by the piracy literature, as piracy may encourage users to experiment and discover (and, hence, reveal) their preference for specific products (Peitz and Waelbroeck (2006); Gans (2012); Zhang (2016)). Indeed, a novel aspect of our experimental design is its use of such information to exogenously reduce user search costs. Our paper also joins a small but growing literature that studies proactive supplier responses that improve the availability and attractiveness of legal options (Danaher et al. (2010); Mortimer et al. (2012); Aguiar and Waldfogel (2018); de Matos et al. (forthcoming)).

## 2 Experimental Design

Table 1 summarizes our experimental design: there are two control and four treatment groups.

We further vary the treatment conditions in two dimensions. 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 for each of the four recommended images. 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.

Cases included in the experiment involve small businesses whose infringement 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.

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).<sup>7</sup> We focus on outcomes in the first 14 days to avoid the confounding effect that worse-performing groups in the initial two weeks received more intense follow-up.<sup>8</sup> 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 includes 7,407 cases; appendix table A4 shows that the groups are well-balanced with respect to basic case characteristics.

## 3 Results

### 3.1 Effects of awareness of the micro-stock licensing option

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

<sup>6</sup>Appendix B provides more details about the experiment and presents the email templates (appendix figures B1-B6). Appendix tables A1 and A2 provide allocation data and balance tests for cases sent.

<sup>7</sup>Appendix table A3 confirms that both likelihoods are statistically the same across groups.

<sup>8</sup>Cases that had not yet licensed received follow-up emails after 14 days.

opened emails, and  $p$ -value  $< 0.001$ ). Almost all of the licenses in the treatment groups are of the micro-stock option (only four licenses are of the premium option)<sup>9</sup>, indicating that the premium option is too expensive for most firms in our sample, and that awareness of an affordable option can significantly increase the licensing rate. Consistent with the idea that infringing use may be informative of demand, the average licensing rate by the four treatment groups is several times higher than the Agency's most effective email marketing campaigns for micro-stock images<sup>10</sup>.

Moreover, within the two control groups, we find that users not given the premium price information are more than four times likely to click the email links to the premium site than users given this information (0.190 vs. 0.045,  $p$ -value  $< 0.001$ ; reported in table 4), suggesting that users' prior beliefs about the premium price are substantially lower than the actual price.

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. We look at each intervention separately, pooling observations from both treatment groups for the other intervention. We discuss detailed break-outs as necessary.

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

Column 1 of table 2a shows that recommending images increases the likelihood of licensing by 45 percent, from 0.020 to 0.029 ( $p$ -value 0.058). Column 2 shows that the average revenue per case increases by 60 percent, from 0.366 to 0.586 ( $p$ -value 0.054). 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 (column 3).

Table 2b shows that providing premium price information increases the licensing rate by 31

### 3.3 Search and potential mechanisms

The above results show that both interventions have an economically large effect on licensing outcomes. We can also 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 conceptual 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) but needs to learn the price. 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, whose quality is uncertain. We assume the user has three choices: (1) search for the price information on the premium site first, and, if the price is too high, decide whether or not to search for a replacement image on the micro-stock site; (2) go to the micro-stock site directly; or (3) take the outside option (which includes, for example,





price information is not given, the difference in the probability of searching the premium site is 2.5 percentage points (p-value is 0.074), which is larger than the 1.1 percentage point difference when users are given the premium price information (p-value 0.169). These results are consistent with an increasing marginal search cost and the notion that image recommendations may mitigate the opportunity cost of learning the premium price <sup>13</sup>.

Finally, we show that, conditional on licensing, half of the images purchased by users who received recommendations are recommended images. In comparison, 56 percent of 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.

#### Provision of the premium-image price information

Conceptually, the effect of the premium-image price information on the probability of visiting the premium site depends on how the actual price compares to the users' prior beliefs about the price level. The comparison between the two control groups, reported in section 3.1, suggests that users' prior beliefs are substantially lower than the actual price level. Column 1 of table 3b confirms this result: within the four treatment groups, the likelihood of searching the premium site after receiving the price information is less than one third the likelihood without receiving this information (0.042 versus 0.129, p-value < 0.001).

Whether or not the premium price information affects users' likelihood of searching the micro-stock site is less obvious. This is because users who would have searched the premium site (in the absence of the price-information intervention) can continue to search the micro-stock site. In our conceptual model, with a constant marginal search cost, the total likelihood of searching the micro-stock site stays the same. However, if the marginal cost increases with additional search, we may observe more search on the micro-stock site by price-informed users because these users, without having to spend time discovering the premium price information, now face a lower cost for searching the micro-stock site.

Consistent with an increasing marginal search cost, column 2 of table 3b shows that providing

---

<sup>13</sup>Note that clicking onto a site does not necessarily constitute a serious search. For example, users may click only to see larger versions of the recommended images, and more email links may simply grab people's attention. These alternative explanations alone, however, cannot explain the overall data patterns.

premium price information increases the likelihood of visiting the micro-stock site by 18 percent (from 0.087 to 0.103, p-value 0.019). Different from the case of image recommendations, we do not observe a lower likelihood of licensing conditional on searching the site (see column 3). This is consistent with the idea that those induced to search the micro-stock site after receiving the premium price information are likely to be infra-marginal users (i.e., those with relatively high willingness to pay), who would have searched for the premium price.

## 4 Conclusion

In this study, we contacted infringing firms of expensive digital images and directed them toward a significantly cheaper product. We designed two interventions to reduce search costs for (i) replacement images and (ii) the price comparison of the two products. We find that awareness of the cheaper option leads to a fourteen-fold increase in the licensing probability. Both interventions have a large positive effect on the licensing rate, and evidence suggests that users face nontrivial, potentially increasing (marginal) search costs. These findings highlight the importance of simplifying the search and transaction process for digital goods and other small-value transactions.

We caution that our results need not reflect an equilibrium effect, as making it easier to license ex-post may affect the demand for ex-ante licensing and, in turn, the overall list prices. Similarly, making the cheaper option easier to use could reduce overall firm profits unless additional changes are made to make the premium products and the process of using them more attractive<sup>14</sup>.

## References

- Aguiar, L. and J. Waldfogel (2018). As streaming reaches food stage, does it stimulate or depress music sales? *International Journal of Industrial Organization* 57, 278{307.
- Anand, B. N. and R. Shachar (2011). Advertising, the matchmaker. *RAND Journal of Economics* 42, 205{245.
- 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.

---

<sup>14</sup>Including the premium licenses, the expected revenue per case for the treatment groups is only 2.5 times that for the control groups (\$0.92 vs. \$0.38; p-value = 0.218), despite the 14-fold increase in the licensing rate. This is because the revenue per premium license is much higher than that for the micro-stock option (\$423 vs. \$23 per license).





Table 3: Average effects of the two interventions on click-through rates onto the licensing sites

(a) Image recommendation					
	Premium site		Micro-stock site		
	N	Browse (1)	Browse (2)	If browse	
				N	License (3)
Without image recommendations	1,762	0.073	0.057	100	0.350
With image recommendations	3,894	0.091	0.112	438	0.253
(p-value)		(0.021)	(0.000)		(0.050)

(b) Premium-image price information					
	Premium site		Micro-stock site		
	N	Browse (1)	Browse (2)	If browse	
				N	License (3)
Without premium-image price information	2,812	0.129	0.087	244	0.258
With premium-image price information	2,844	0.042	0.103	294	0.282
(p-value)		(0.000)	(0.033)		(0.532)

Notes

Online Appendices (not for publication)



Table A3: Generating the analysis sample

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

Notes: We remove 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.



Table A4: Balance tests for cases in the analysis sample

	N	Multi-image case	High resolution	Secondary pages	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 (0.31)	0.67 (0.76)	0.79 (0.12)	14.46 (0.43)	1.49 (0.82)
Micro	865	0.13 (0.92)	0.67 (0.57)	0.80 (0.14)	14.80 (0.39)	1.41 (0.22)
Micro + Rec	1,947	0.11 (0.02)	0.68 (0.91)	0.80 (0.17)	14.50 (0.49)	1.51 (0.93)
Micro + Price	897	0.10 (0.03)	0.68 (0.87)	0.80 (0.14)	14.48 (0.48)	1.53 (0.68)
Micro + Rec + Price	1,947	0.12 (0.13)	0.68 (0.96)	0.81 (0.50)	14.44 (0.32)	1.56 (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;

Table A5: Licensing and click-through rates of the micro-stock option: the four treatment groups

(a) Licensing			
	Without premium-image price information	With premium-image price information	(p-value)
Without image recommendations	0.015	0.025	(0.153)
With image recommendations	0.026	0.031	(0.290)
(p-value)	(0.078)	(0.317)	

(b) Click-through rates			
	Without premium-image price information	With premium-image price information	(p-value)
Without image recommendations	0.037	0.076	(0.000)
With image recommendations	0.109	0.116	(0.478)
(p-value)	(0.000)	(0.001)	

Notes: 14-day licensing and click-through rates of the a ordable, micro-stock option by the four treatment groups in the analysis sample. p-values in parentheses are based on two-sided t-tests.



Table A7: Effects of the two interventions on clicking onto the licensing sites: regression results

	Premium site	Micro-stock site	
	Browse (1)	Browse (2)	If browse License (3)
With image recommendations	0.017** (0.008)	0.055*** (0.007)	-0.088* (0.053)
With premium-image price info	-0.088*** (0.007)	0.016** (0.008)	0.013 (0.039)
log(total stock images + 1)	0.009*** (0.002)	0.012*** (0.003)	0.016 (0.013)
Case age	-0.002*** (0.001)	-0.001 (0.001)	-0.003 (0.005)
Multi-image case	-0.011 (0.011)	-0.020* (0.012)	0.141* (0.074)
High resolution	0.024*** (0.008)	0.011 (0.008)	-0.081* (0.044)
Secondary page	-0.025** (0.010)	-0.019* (0.010)	0.052 (0.045)
Batch fixed effects	Y	Y	Y
N	5,656	5,656	538

Notes

## B. Further details of the experiment

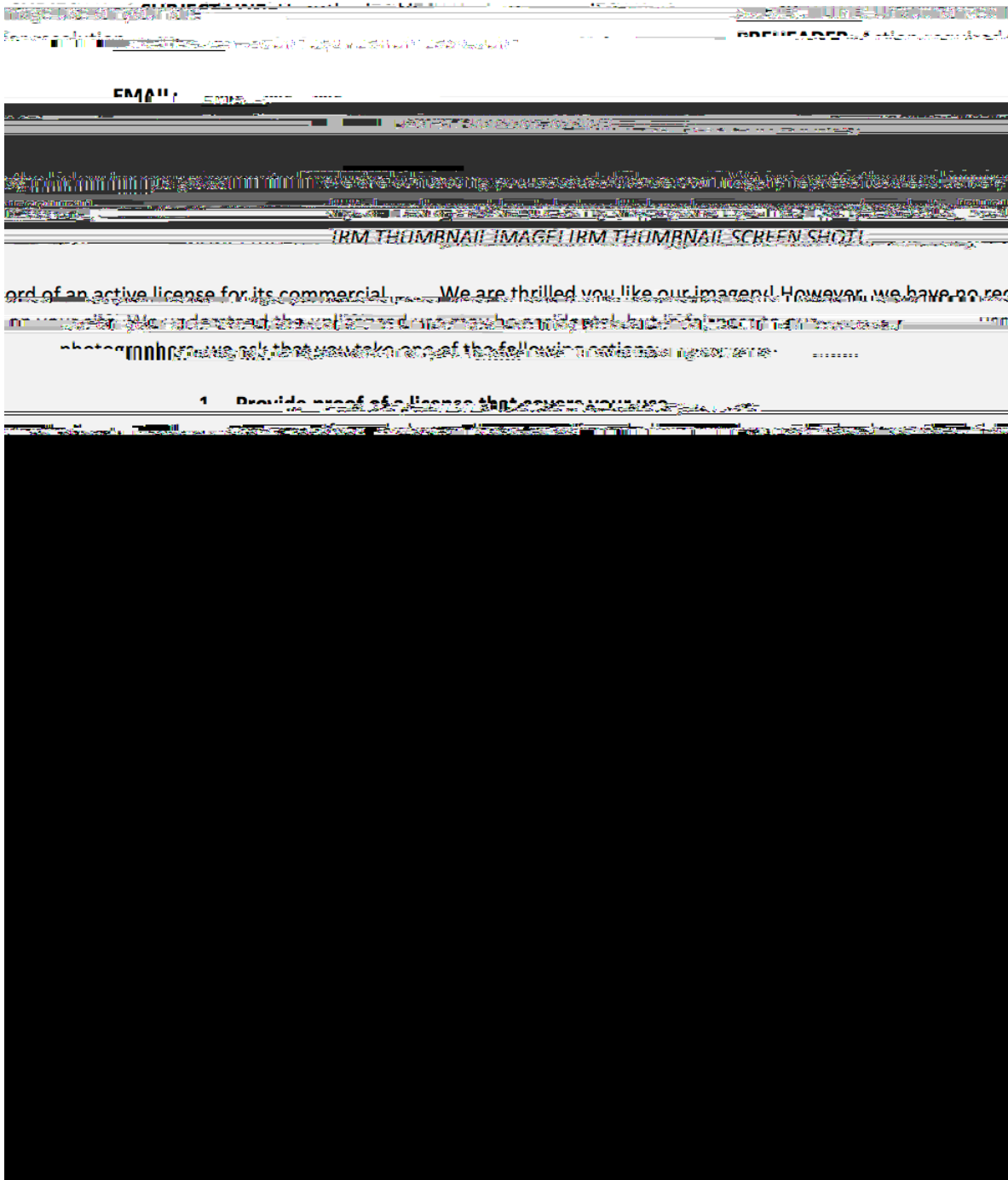
### B.1 Email templates

Figure B1. Template illustration|emails to group \0"



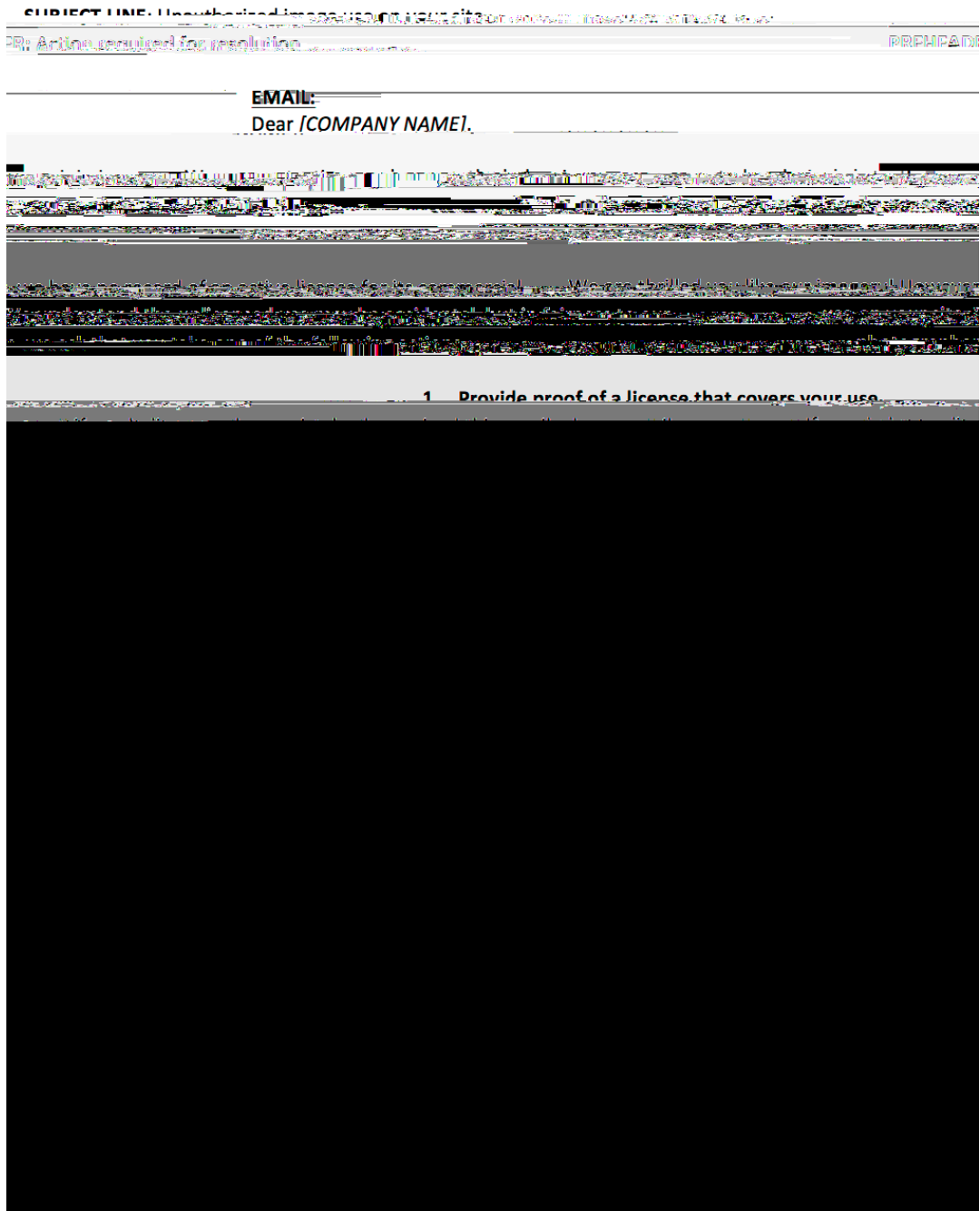
Notes: control group, without premium-image price information.

Figure B2. Template illustration|emails to group \Micro"



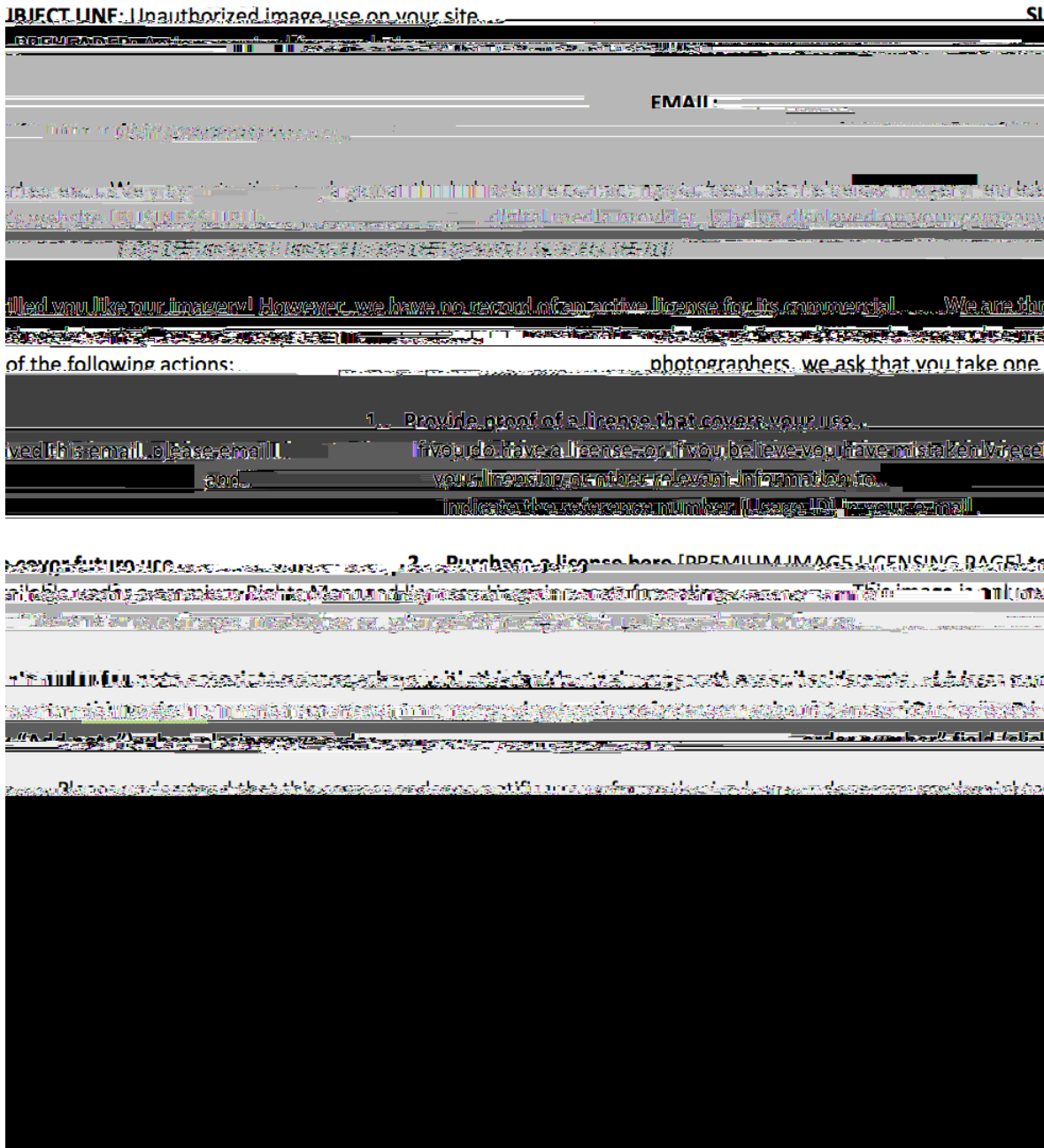
Notes: treatment group, without image recommendation or premium-image price information.

Figure B3. Template illustration|emails to group \Micro + Rec"



Notes: treatment group, 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.

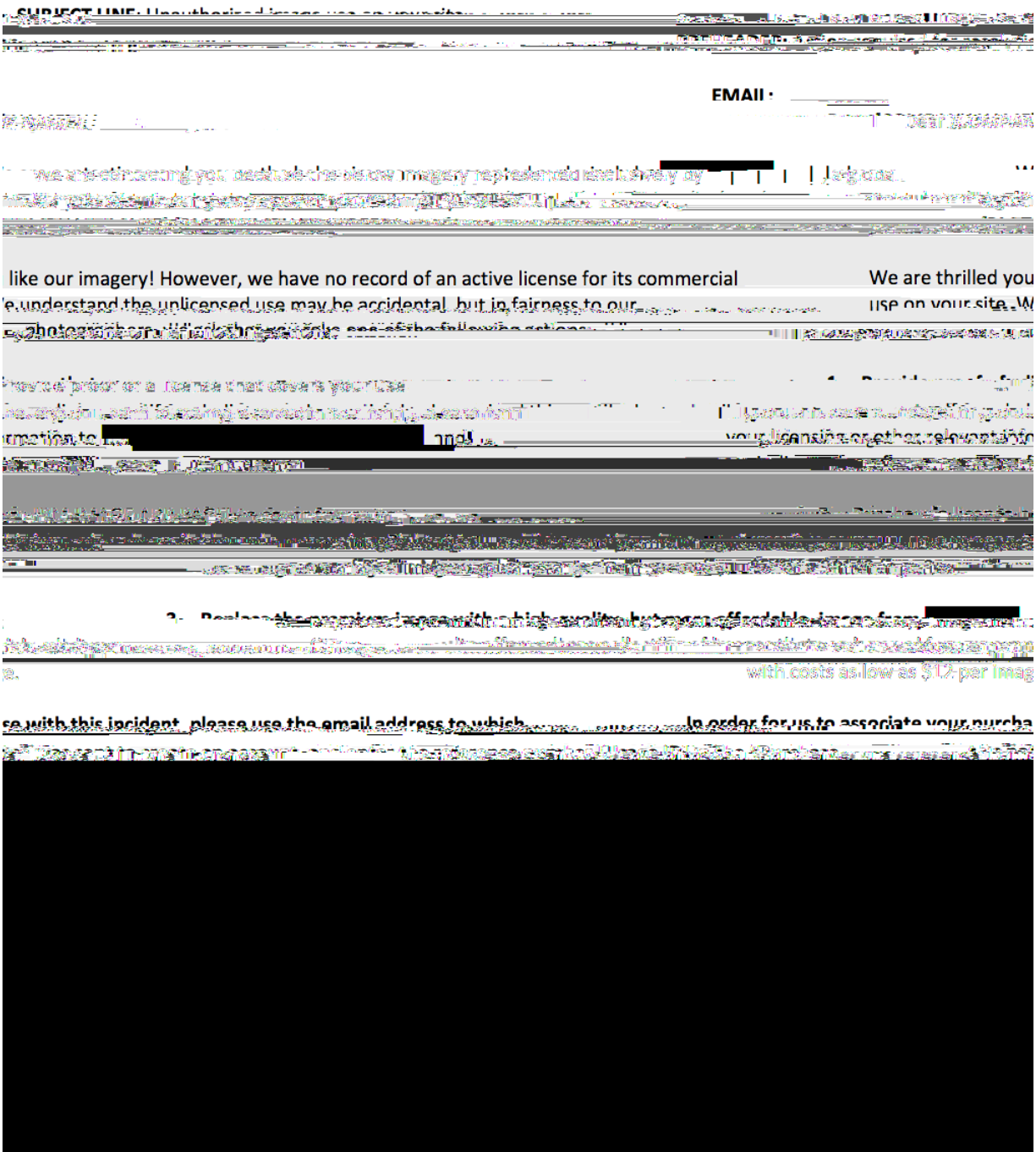
Figure B4. Template illustration | emails to group "Price"



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



Figure B5. Template illustration|emails to group \Micro + Price"



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

Figure B6. Template illustration|emails to group \Micro + Rec + Price"

**SUBJECT LINE:** Unauthorized image use on your site



Notes

## 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, which the Agency deemed to be the most constructive approach. It turned out later that the similar-image algorithm had not worked initially 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 randomly added these cases equally to the four groups that were not given image recommendations.

We planned to send 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 cases that the Agency sent out was 24,090; and, 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 their 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