Product Lineups: The More You Search, The Less You Find

SANG KYU PARK
ANER SELA

Consumers often try to visually identify a previously encountered product among a sequence of similar items, guided only by their memory and a few general search terms. What determines their success at correctly identifying the target product in such "product lineups"? The current research finds that the longer consumers search sequentially, the more conservative and—ironically—inaccurate judges they become. Consequently, the more consumers search, the more likely they are to erroneously reject the correct target when it finally appears in the lineup. This happens because each time consumers evaluate a similar item in the lineup, and determine that it is not the option for which they have been looking, they draw an implicit inference that the correct target should feel more familiar than the similar items rejected up to that point. This causes the subjective feeling of familiarity consumers expect to experience with the true target to progressively escalate, making them more conservative but also less accurate judges. The findings have practical implications for consumers and marketers, and make theoretical contributions to research on inference-making, online search, and product recognition.

Keywords: lineups, product search, product recognition, familiarity judgments, signal detection

One of the authors recently visited an acquaintance's house, where he sat in an exceptionally comfortable chair, offering unmatched fit and support. After the visit, the author tried to find the exact same model on the manufacturer's website, but found himself browsing through dozens of similar-looking chairs, guided only by a fuzzy recollection of the shape and size of the chair in which he sat earlier. Such product search experiences appear quite common. Indeed, a survey we conducted suggests that a majority of consumers have occasionally tried to visually identify a previously encountered product among numerous similar options examined sequentially (see web appendix A). In such cases, what influences consumers' likelihood of correctly identifying the target product when they finally encounter it in their search?

Despite the everyday prevalence of such experiences—which we label "product lineups"—little prior research has examined them directly. Previous work has examined how memory impacts product recognition (Bettman 1979), how advertisements influence recall (Norris and Colman 1992), how the size of the set under consideration influences decision quality and satisfaction (Diehl 2005; Diehl and Poynor 2010), and factors that affect how long consumers search for the lowest price (Rowley 2000; Stigler 1961; Titus and Everett 1995). Little is known, however, about how the internal dynamics of product search experiences may influence the accuracy with which consumers are able to identify a previously encountered target product embedded in a sequence of relevant but incorrect options.

We examine how identification accuracy is influenced by the dynamics of the sequential search experience itself. Specifically, we propose that as consumers screen options to find the product they encountered before, their internal matching threshold (i.e., the level of perceived fit beyond

Sang Kyu Park (sangkyu.park@warrington.ufl.edu) is a doctoral student in marketing, and Aner Sela (aner.sela@warrington.ufl.edu) is an associate professor of marketing, at the University of Florida, 267 Stuzin Hall, Gainesville, FL 32611. Please address correspondence to Sang Kyu Park. This article is based on the first author's dissertation, chaired by the second author. Supplementary materials can be found in the web appendix accompanying the online version of this article.

Editor: Amna Kirmani

Associate Editor: Cait Lamberton

Advance Access publication January 9, 2020

 • Vol. 47 • 2020
DOI: 10.1093/jcr/ucaa001

which they conclude that they have found the option for which they have been looking) becomes progressively higher. This elevated matching threshold leads consumers to be more likely to falsely reject the true target for which they have been looking when they finally encounter it (i.e., to misidentify the correct product as incorrect). Consequently, screening more options makes consumers more conservative judges, but—ironically—it may also make them less likely to correctly identify the true target.

We hypothesize that consumers' matching threshold may escalate because each time consumers evaluate a relevant option in a product lineup and judge it to be incorrect, they draw an implicit inference that the true target must feel even more familiar or right than the options screened up to that point. That is, people's implicit expectation for the level of felt fit between the correct target and the representation they hold in their memory is ratcheted up each time they dismiss an option that does not feel sufficiently right.

This research makes several contributions. First, it introduces the notion of product lineups and examines a novel process that influences the accuracy with which consumers are able to identify options. Second, it extends prior research on visual recognition, specifically as it applies to recognizing a previously seen item among multiple lookalikes. Third, the findings advance understanding of mechanism that may bias respondents' identification accuracy in other contexts beyond product identification, such as police lineups (see Reisberg 2014 for a review).

THEORETICAL BACKGROUND

Consumers often visually search for a product they encountered previously without knowing its exact name. Whether we see a product in an advertisement but fail to register the name of the brand, observe others using an attractive product but are reluctant or unable to ask for details, or directly experience a high-quality product, we often later find ourselves looking for the specific product we encountered before. Consumers often wish to find a specific product they saw earlier, rather than an alternative, because they believe that particular option would deliver high quality or performance (e.g., athletic gear or technical equipment), convey a social signal or another symbolic trait (e.g., a fashion statement, fitting in), or satisfy their curiosity regarding the previously encountered item, even if they are open to buying something else eventually. In all these scenarios, consumers may try to visually identify a specific brand, model, or product among many lookalikes (to which we refer here as "lures"), guided only by a fuzzy recollection of the previously encountered target.

Screening options in a sequential product lineup is, by definition, a product search task. However, it differs from the prototypical task examined in many classic works on consumer search (Hoch, Bradlow, and Wansink 1999; Moorthy, Ratchford, and Talukdar 1997; Rothschild 1978; Stigler 1961; Zwick et al. 2003). Such work has typically focused on how long the consumer continues to search, given a search cost, until finding an option that is good enough (e.g., meets the consumer's reservation price, which is often unknown and determined on the fly). In contrast, we examine situations where consumers try to assess the visual match between the options they encounter in the search and a representation they hold in memory of a *specific* item encountered previously (i.e., "is this what I have seen before?"). Further, although consumers' preferences or ideal points may sometimes shift during search, as consumers discover new features or options (Hauser, Dong, and Ding 2014), preferences are less likely to shift in the situations we described above, where consumers strive to identify a specific previously encountered product, regardless of how attractive other items in the lineup appear to be. Thus, the current research holds preference constant during the search process and focuses on identification accuracy.

This makes product lineups conceptually similar to visual recognition problems in which respondents are asked to determine whether a stimulus (e.g., an image) presented to them has or has not been previously seen (Benjamin and Bawa 2004; Meissner et al. 2005; Palmer and Brewer 2012; Segal and Fusella, 1970). Visual recognition research has often used signal detection theory as a conceptual framework for understanding visual recognition effects. We briefly review some of the basic assumptions of this theory, which we use as a basis for developing our predictions.

According to signal detection models, any given stimulus can be thought of as positioned along a continuum of familiarity or strength of evidence, where previously encountered items are positioned higher than lures or distractors not previously encountered (Green and Swets 1966; Palmer and Brewer 2012; Pleskac 2007; Snodgrass, Volvovitz, and Walfish 1972; Wixted 2007). These models suggest that people base their identification judgments on some internal criterion, or threshold, for the strength of evidence they require (i.e., the degree of perceived match). A stimulus is judged to have been seen before if its level of perceived match exceeds the internal decision threshold. Novel items typically give rise to lower levels of perceived match, which increases the likelihood that they be judged as novel, whereas the opposite is true for previously encountered items. This basic notion of threshold-based judgment is also a central tenet of global memory models (Ashby and Perrin 1988; Eich 1982; Gillund and Shiffrin 1984; Nosofsky and Zaki 2003), which assume that any stimulus gives rise to some level of felt familiarity when encountered by a respondent. This perceived degree of familiarity is then compared to an internal threshold, to arrive at a conclusion (e.g., "this feels like something I've seen before").

The identification decision thus depends on where the threshold is located. A stringent threshold implies that a higher level of perceived match is required to elicit a *yes* response (Palmer and Brewer 2012), which decreases the likelihood of false identification but also that of correctly identifying the true target. A lax threshold increases the chances of false identification but also that of correctly identifying the true target (i.e., anything that looks reasonably similar to the stored representation will be judged as a match). Note that threshold location does not necessarily entail, in and of itself, a change in discriminability or decision accuracy (Palmer and Brewer 2012).

A well-established finding is that threshold placement is malleable (Brown, Lewis, and Monk 1977; Hirshman 1995). Matching threshold often shifts as a result of cues that are independent of the evaluation process, such as providing respondents with accuracy feedback (Rhodes and Jacoby 2007), instructions to use a more stringent or more lax threshold (Estes and Maddox 1995), time delay between target encoding and the identification task (Singer, Gagnon, and Richards 2002; Singer and Wixted 2006), the perceived difficulty of recognition (Benjamin and Bawa 2004), and people's beliefs regarding their own memory accuracy (Miller and Wolford 1999; Verde and Rotello 2007). Whereas prior research has mostly focused on external cues that affect this threshold, we explore how the internal dynamics of the task (i.e., having already rejected incorrect items) may also influence the threshold placement.

THRESHOLD ESCALATION THROUGH INFERENCE FROM PRIOR REJECTIONS

We propose that consumers' internal matching threshold may progressively shift as a result of mismatch judgments on prior items (i.e., lures). In particular, we propose a novel mechanism where the more decision-makers reject similar but incorrect lures, the more stringent their matching threshold becomes. Consequently, decision-makers become increasingly likely to erroneously reject the correct target when it is finally encountered.

We argue that each time decision-makers evaluate whether an option in a product lineup is the product seen earlier, and judge it to be different (i.e., falling below their internal matching threshold), they come to expect the correct target to provide a greater feeling of fit or rightness (Cesario, Grant, and Higgins 2004; Schwarz 2006). That is, people may reasonably infer that the correct target, when encountered, should feel significantly more familiar than any incorrect lure. Our prediction is consistent with the notion that judgments of a target often shift disproportionately in a direction opposite to a comparison standard from which an item is perceived as different (Ford and Thompson 2000; Mussweiler 2003; Sherif and Hovland 1961; Shoots-Reinhard et al. 2014; Zellner et al. 2003). For example, liking of a positive stimulus is amplified following a less positive stimulus (Dolese et al. 2005; Parker et al. 2008); and after consumers view a moving target, a new target must move even faster in the same direction to appear to be in motion (Warren 1985). Extending this logic, we predict that after consumers judge a similar lure in a product lineup as incorrect, they may expect the correct target to feel even more familiar, by way of contrast. Consequently, after rejecting a lure, their internal matching threshold should escalate. Namely, subsequent candidates have to feel even more right to be judged as correct.

We further propose that this inference is inflated each time another lure is evaluated in a sequence. Prior research underscores the effect of repetition on inflated perceptions and evaluations. For example, mere repetition of attitude expression induces attitude polarization (Downing, Judd, and Brauer 1992; Fazio et al. 1982) and inflates attitude conviction and certainty (Holland, Verplanken, and van Knippenberg 2003). Similarly, thinking more about an object leads to more extreme evaluations (Tesser and Leone 1977; Tesser, Martin, and Mendolia 1995). This inflation-through-repetition principle has been observed both for nonevaluative and evaluative judgments (see Judd and Brauer 1995 for a review).

Because judging and dismissing lures in a product lineup is inherently reiterative, we predict that rejecting a longer sequence of lures will result in more pronounced threshold escalation, compared with a shorter sequence. Thus, the more decision-makers reject similar but incorrect items in a product lineup, the higher their internal matching threshold becomes.

How might such ratcheting up of the internal threshold impact identification accuracy? Ironically, although threshold escalation makes decision-makers more conservative, it may also decrease their identification accuracy because it requires a stronger feeling of familiarity for an item to be considered the correct target (Palmer and Brewer 2012). Perceptual memory is often fuzzy and noisy (Krueger 1978; Schacter 1999), particularly for targets that are not highly familiar (Hockley 2008; Megreya and Burton 2007; Stretch and Wixted 1998). The stored representation of the original target may also be distorted due to various perceptual or attitudinal factors (Beckwith and Lehmann 1975; Huber and Holbrook 1979). Consequently, the benchmark upon which the matching threshold is based is itself unstable and largely constructed (Whittlesea, Jacoby, and Girard 1990). It is therefore possible for the escalated matching threshold to surpass even the degree of perceived match actually offered by the true target, resulting in the true target being mistakenly judged as previously unseen when the internal threshold is sufficiently high. Much like a type II error in null hypothesis testing, a more stringent threshold does not necessarily imply that the decision-maker becomes more discerning of the correct target; rather, it implies that the decision-maker will require stronger evidence (i.e., a feeling of match or familiarity) to conclude that the item in question is the correct target.

In summary, we propose that, as consumers screen more options in a product lineup, their internal threshold for judging an item as the one they are seeking increases. This makes consumers more likely to erroneously misidentify the correct target as incorrect, the more they have already screened similar but incorrect lures.

Note that our account focuses on shifts in the location of the internal threshold decision-makers use to make a match judgment, rather than the fluency of the target item (Whittlesea et al. 1990), the activation of a rejection mindset or momentum (Xu and Wyer, 2012), fatigue, memory decay, or inaction inertia (where people decline inferior options after rejecting superior ones; Tykocinski and Pittman 1998)—all of which are inconsistent with our findings. We also test whether the effects may be driven by a naïve Bayesian inference about the chance of encountering the correct target, decreased confidence in one's judgment, or doubt that the correct target will ever appear in the lineup.

Five studies, involving incentive-compatible as well as hypothetical decisions, provide novel insights into the dynamic nature of product lineups and demonstrate how threshold escalation influences product identification. Studies 1 and 2 demonstrate the effect while ruling out several alternative accounts, such as rejection momentum or inertia, fatigue, memory decay, distorted base-rate perception, participant attrition, doubt, and mere error. Studies 3 and 4 use a moderation approach to demonstrate the inferential process underlying the effect, while further casting doubt on alternative explanations based on rejection inertia, doubt, memory decay, experimental demand, and mere error. Study 5 identifies a boundary condition where consecutive mismatch judgments do not influence correct target identification accuracy. In all our studies, we a priori exclude participants who fail standard attention checks (Oppenheimer, Meyvis, and Davidenko 2009). We conclude with a discussion of the theoretical and practical implications. Our web appendix includes additional studies examining the effect of similarity between the target and lures and the effect of lineup length on identification accuracy.

STUDY 1: THRESHOLD ESCALATION IN PRODUCT LINEUPS

Study 1 provides preliminary evidence for the threshold escalation effect using a naturalistic, incentive-compatible design. Participants first formed, through direct experience, an authentic individual preference for a product. Then, we incentivized them to find their preferred product in a lineup, among several similar-looking but allegedly lower-quality lures, by giving them a chance to receive their chosen option. We predicted that participants would be more likely to falsely judge the correct target as incorrect when the target was preceded by a longer sequence of relevant but incorrect lures, due to an increase in participants' internal matching threshold.

Method

One hundred six participants (M_{age} = 20.7, 45.3% women; no participants excluded from analyses) were individually approached and recruited on the campus of a large university. We collected responses across several days, and the day on which the data was collected had no effect on the results. We told participants that they would be participating in a marketing study.

Respondents who agreed to participate received two different Logitech wireless computer mouse devices. An experimenter blind to the experimental condition verbally described the two devices as having both received the highest objective ratings among all other similar products on the market and as having the longest battery life, highest precision, and strongest connectivity. Participants were then asked to physically examine the two devices and select their preferred option (66.3% preferred one mouse, but preference did not influence the results nor interact with any of our factors—all $ps > .3$—so it is not discussed further). We asked them to briefly explain why they preferred one mouse over the other, and then handed them a tablet computer on which they completed the remainder of the study. When they received the tablet, we unobtrusively took back the two devices and put them inside an opaque bag so that they could no longer be seen.

On the tablet, participants indicated which of the two options they chose and rated how comfortable the selected mouse felt and how likely they were to buy it. They also indicated whether they had seen or used any of the two devices before. Twenty-five participants responded affirmatively, and excluding them had no effect on the results. Next, participants watched a short advertisement clip, which served to cleanse their working memory. We then told participants that they had a chance to win their preferred mouse, as a token of gratitude, if they could identify it among other options. Participants were told they would see several mouse devices (we did not tell them exactly how many) and that they could choose the one they would like to win. We told them that the devices they saw at the beginning of the study had received higher quality ratings, so it was in their best interest to find the same product they preferred earlier.

At this point, participants were randomly assigned to either a two-lures or a six-lures condition. Participants in the six-lures condition saw six lure devices, visually similar but not identical to the target, in a random order (see appendix A). Each lure appeared on a separate page along with a fictitious product number (e.g., Logitech #AN3) and a binary-choice question asking participants whether that was the product they would choose for a chance to win (yes or no). No additional product attribute information was provided, reflecting the notion that this was a visual recognition task (similar to when consumers examine product pictures online) rather than a new preference formation task. Those in the two-lures condition saw two lure products randomly drawn from among the six options used in the six-lures condition.

Finally, after seeing two or six lure mouse options, all participants saw the mouse they had claimed earlier to be their preferred option (i.e., the correct target), followed by the same binary identification question. The presentation of the correct target was seamless in that the format and question wording were identical to the lures. Identification accuracy of the correct target was our dependent variable.

Participants who selected a mouse other than their previously preferred option were later asked whether they did so because they found one of the lures more attractive or because they thought it was the correct target. Eight participants indicated that they chose a lure device because they found it more attractive, and these participants were excluded from the analysis because they deviated from our instructions to search for the correct target (including these participants does not change the results). We randomly selected two winners, who received their preferred mouse two weeks after data collection was completed.

Results

The Threshold Escalation Effect. We proposed that each time people judge a similar but not identical option as a mismatch, their matching threshold escalates. Consequently, the more mismatch judgments they make before encountering the correct target, the more likely they are to erroneously judge the target as incorrect when they finally see it. Consistent with our hypothesis, participants were less likely to correctly identify the target product (i.e., they were more likely to judge it as incorrect) after seeing six lures (39.1%) than after seeing only two lures (65.4%; $\chi^2(1) = 6.76$, $p = .009$, Cohen's $d = .54$).

Overall, similar but incorrect devices (i.e., lures) were correctly rejected 81.6% of the time. Because there were more lures in the six-lure than in the two-lure condition, one might wonder whether the results are driven by a greater error rate in the six-lure condition. For example, participants who are innately less conservative (i.e., have a lower matching threshold to begin with) might have been more likely to misidentify a lure as the correct target in the six-lure condition, and these participants might have later been reluctant to identify the true target as correct because they wished to appear consistent. This may increase the rate of false rejections of the true target in the six-lure condition. Of note, however, the percentage of participants who mistakenly responded "yes" to a lure was not significantly different across conditions, ($\chi^2(1) = 1.61$, $p = .20$). Moreover, analyzing only participants who correctly rejected all the lures ($n = 45$) presented to them reveals a slightly stronger (rather than weaker) effect on false rejection of the true target (six lures: 38.9% vs. two lures: 81.5%; $\chi^2(1) = 8.55$, $p = .003$, Cohen's $d = .97$). Taken together, these findings are inconsistent with the possibility that the effect was driven by unequal proportions of participants who misidentified a lure as the correct target.

Our threshold escalation account receives direct support from examination of judgment accuracy for the last lure, presented right before the correct target (i.e., the sixth lure in the six-lure condition and the second lure in the two-lure condition). If the effect is driven by threshold escalation, as we argue (i.e., increased conservativeness after seeing six rather than two lures), then participants should be more likely to correctly identify the sixth lure as such, compared with the second lure. This should be true as long as the feeling of fit provided by the lures is sufficiently close to (e.g., just under) participants' baseline matching threshold. When this is the case, moving the threshold upward should decrease the likelihood that participants mistake the lure for the target. Increased conservativeness may not necessarily increase judgment accuracy for lures that are very easy to discern (i.e., provide a feeling of fit well below participants' baseline threshold). In such cases, a moderate increase in conservativeness may not significantly benefit judgment accuracy.

Consistent with a threshold escalation mechanism, participants were more likely to correctly identify the last lure in the sequence in the six-lure condition (84.8%) than in the two-lure condition (65.4%; $\chi^2(1) = 4.83$, $p = .028$, Cohen's $d = .46$). Similarly, within the six-lure condition, participants were more likely to correctly identify the sixth lure (84.8%) than the second lure (65.2%; $\chi^2(1) = 4.70$, $p = .03$). Recognition accuracy was not different for the second lure across conditions ($\chi^2(1) = .4$, $p = .53$, Cohen's $d = .45$). These results indicate that our effect is driven by a higher matching threshold (i.e., a tendency to be more conservative in judgments of familiarity) in the six-lure condition compared with the two-lure condition. Identification rates for each sequence of judgments are reported in web appendix B.

Mere Inaccuracy?. One may wonder whether the higher false rejection rate in the six-lures condition might have been caused by confusion, depletion, or memory decay after consumers viewed a larger number of lures, resulting in increased inaccuracy. However, note that such alternative explanations would predict a higher false-negative error rate for the correct target *as well as* a higher false-positive error rate for the lure preceding the target, which is opposite to what we found.

Follow-Up Study: Testing a Bayesian Updating Alternative Account

An alternative explanation for our finding is that seeing more lures may lead participants to infer that there are simply more lures in the market, so the true target is less likely to appear on any given trial, in line with the notion of naïve Bayesian updating. This, rather than the threshold escalation mechanism we propose, may lead people to become more conservative judges after seeing a longer sequence of lures. To test this alternative account, we conceptually replicated study 1, but told participants ($N = 158$) there were exactly 15 possible products in that market, thereby holding base-rate perceptions constant across conditions (see web appendix C). The results were identical to study 1: screening a greater number of lures increased the likelihood of erroneously rejecting the correct target even when participants received a concrete base-rate benchmark (six lures: 61.3% vs. two lures: 82.1%, $\chi^2(1) = 8.39$, $p = .004$, Cohen's $d = .46$). The effect was robust when only those who correctly rejected all the lures were included ($n = 128$, six lures: 65.6% vs. two lures: 85.1%, $\chi^2(1) = 6.62$, $p = .01$, Cohen's $d = .47$).

Discussion

Study 1 demonstrates the threshold escalation effect using a naturalistic, incentive-compatible design. Participants formed a genuine preference for a product through direct physical experience. They then tried to visually identify their preferred product among several similar but inferior lures, as a part of a product lineup, and they had a chance to actually receive the product they chose. Thus, their decision had a real consequence. The follow-up study demonstrates the generalizability of the effect while also ruling out an alternative explanation based on base-rate updating.

The findings were consistent with our matching threshold escalation hypothesis. First, identification accuracy for the correct target *decreased* (i.e., the target was more likely to be judged as incorrect) after participants screened six rather than two lures. Further bolstering the mechanism, participants' identification accuracy for incorrect lures *increased* after they screened six rather than two lures. This finding is consistent with the notion that participants become more conservative in their judgments (rather than merely less accurate) the more lures they screen.

The findings rule out alternative explanations based on confusion, depletion, and memory decay. Such alternative accounts would predict an overall increase in error rate—for both the correct target *and* incorrect lures—when more lures are screened.

A limitation of study 1 is that its realism came at the cost of allowing participants to occasionally misidentify lures as the true target. This gives rise to endogeneity concerns—namely, that our results might be biased by participants who incorrectly misidentified a lure as the true target. Although the results are unaffected by whether these participants are included in the analysis, a more ideal procedure would prevent participants from erroneously identifying lures as the correct target. We develop such a procedure in study 2.

STUDY 2: THRESHOLD ESCALATION IN A CONTROLLED SETTING

Study 2 replicates the findings in our previous studies in a more controlled setting, using simulated mismatch judgments for the lures. As in the previous study, we showed participants several lures similar to the target product, but we eliminated the possibility of participants identifying one of the lures as the target.

Method

Two hundred ninety-three undergraduate students were recruited form a large university ($M_{age} = 20.1$, 52.6% women; excluding 14 participants who did not follow the instructions; including them in the analyses did not significantly alter the results). They were randomly assigned to either a five- or nine-lures condition, instead of two versus six in the previous study. We varied the number of lures for generalizability.

Participants saw a picture of a pair of headphones described to be of high quality and were asked to imagine that they saw someone wearing it recently. Participants further imagined that they wanted to search online for the exact product they had seen, and that they were unwilling to settle for similar options (similar to the follow-up study above; web appendix C). This instruction was provided after the image of the target product disappeared, to minimize attempts to deliberately memorize the image.

Next, to cleanse their working memory, we had participants in both conditions complete a filler task from previous research (i.e., reading a short essay on dolphins; Sela and Shiv 2009).

We then told participants that they were going to see a series of headphones. We reminded them that they were going to browse through the options as if they were shopping online, and that their goal was to identify the exact pair of headphones they had seen earlier.

Similar to the previous study, those in the nine-lures (five-lures) condition saw nine (five) lure headphones (web appendix D) in a random order, each on a separate page. Each time participants saw a lure, we asked them to examine the product and decide whether that item was the correct target, but we did not ask participants to enter their response. Instead, after 5 seconds, we told participants that "although it seems pretty similar, you realize that this is not the product you were looking for" (web appendix D). This procedure simulates the same mismatch judgment while preventing participants from incorrectly identifying one of the lures as the target. Prior research shows that simulating an information processing experience is often sufficient to elicit the effects usually observed with the actual experience (Novemsky et al. 2007; Wänke, Bohner, and Jurkowitsch 1997), and mentally simulating a perceptual judgment can influence the evaluation of that judgment in subsequent inferences (Winkielman and Schwarz 2001). Note that, for the purpose of testing our theory, asking participants to mentally simulate rejecting a lure should have the same effect as their spontaneous judgment. This is because our theory focuses on the inference participants draw from the conclusion that the item just seen is not the true target (i.e., inferring that the true target must feel even more right), not necessarily the real-time experience of evaluating the lure itself (e.g., how easy it is to process or identify). Thus, even if participants' spontaneous reaction to the lure was hesitant, assuring them that it was not the correct target should have the same effect on the matching threshold, leading them to expect that the true target should appear more familiar in contrast.

After the lures came the correct target they saw at the beginning of the survey. Participants indicated whether it was the correct product, which was the dependent measure.

To keep the total number of judgments the same across conditions (i.e., nine, not including the target), the first four judgments in the five-lures condition were fillers unrelated to headphones (e.g., "On a scale of 1–7, how much do you like this BeanieBoo?", "How much do you consider yourself a coffee person?"; web appendix D). These were followed by five lure headphones, thereby keeping the total length of the sequence the same across conditions. We did not tell participants in advance how many items they would see.

Results

The results of previous studies were replicated such that participants were less likely to correctly identify the target after screening nine lures (55.5%) than five lures (81.0%; $\chi^2(1) = 21.94$, $p < .001$, Cohen's $d = .57$).

Discussion

Study 2 bolsters the threshold escalation hypothesis and demonstrates that the effect holds regardless of whether participants rejected the lures spontaneously, as in study 1, or mentally simulated it as in study 2: seeing a longer sequence of lures—and recognizing that none was the correct target—increased the tendency to erroneously reject the correct target when it finally appeared.

One may wonder if our results reflect decreased confidence in their ability to make correct judgments as the search continues. We ran two high-powered follow-up studies (combined $N = 736$) to directly test this alternative account (see web appendix E for details) by directly measuring participants' confidence in their judgment after they made their identification decision for the correct target (1 = not at all confident, 9 = very confident). The threshold escalation effect was replicated in both studies (six lures: 73.5% vs. two lures: 83.2%; $\chi^2(1) = 5.54$, $p = .019$, Cohen's $d = .24$; nine lures: 59.8% vs. five lures: 77.5%; $\chi^2(1) = 12.51$, $p < .01$, Cohen's $d = .39$, respectively), but no difference was found for confidence ($M_{\text{six lures}} = 7.07$ vs. $M_{\text{two lures}} = 7.16$; $F(1, 391) = 0.3$, $p = .586$; $M_{\text{nine lures}} = 6.68$ vs. $M_{\text{five lures}} = 6.99$; $F(1, 341) = 2.47$, $p = .12$, respectively). The effect was robust when we controlled for confidence ($p = .019$, .001 respectively), and confidence did not interact with the number of lures screened ($ps > .40$).

In addition, another replication of study 2 ($N = 431$) tested whether showing participants more lures led them to doubt the likelihood that the correct target would reappear, which could also explain the higher false rejection rates (see web appendix E for details). After screening five or nine lures and simulating making mismatch judgments for all of them, as in study 2, participants were asked "How likely are you to encounter the correct product soon?" (1 = very unlikely, 7 = very likely). The results of this high-powered study reveal no effect of condition on doubt ($M_{\text{nine lures}} = 4.26$ vs. $M_{\text{five lures}} = 4.38$; $F(1, 429) = 0.76$, $p = .385$).

Taken together, these findings cast doubt on the possibility that decreased confidence, or doubt in one's judgment or in the likelihood of finding the correct target, is driving our results. Note also that the threshold escalation account does not necessarily predict an increase in confidence when the threshold is higher, consistent with the findings above. Indeed, decision-makers may adjust their threshold to a level that makes them comfortably confident, maintaining a confidence homeostasis of sorts, in much the same way that they adjust their level of deliberation effort to be compatible with task characteristics (Schrift, Netzer, and Kivetz 2011).

STUDY 3: MODERATION EVIDENCE FOR THRESHOLD ESCALATION

Studies 3 and 4 use the moderation-of-process approach to provide evidence for the underlying process (Bullock, Green, and Ha 2010; Spencer, Zanna, and Fong 2005). We directly manipulated the extent to which participants perceived that they should ratchet up their matching threshold after rejecting similar but incorrect lures.

In study 3, we used a manipulation designed to attenuate this perception. Specifically, we used a social comparison manipulation to imply that participants already had a relatively stringent matching threshold, compared with most other participants. We expected this cue to attenuate participants' perception that they should further ratchet up their threshold, thereby attenuating the effect of few versus many lures on target identification. In the control condition, we used a similar social comparison manipulation to imply that participants were doing about the same as most other participants, which provides no new diagnostic information. Thus, we expected the effect of few versus many lures to hold in the control condition, as in our prior studies.

Method

The procedure was similar to that used in study 2 except that we introduced a social comparison manipulation. Three hundred twenty participants from Amazon Mechanical Turk (MTurk) (M_{age} = 37.8, 55.3% women, excluding 20 participants who did not follow the instructions; including them in the analyses did not significantly alter the results) were randomly assigned to a 2 (lures: five vs. nine) × 2 (social comparison cue: lax vs. control) between-subjects design. As in study 2, participants tried to identify a previously presented set of headphones in a product lineup. We used the same procedure described in study 2, where participants mentally simulated making a series of mismatch judgment for either five or nine lures.

Our theory predicts that rejecting more lures (in this case, through simulation) should increase misidentification of the true target at baseline, but not when an external cue led participants to believe that their matching threshold was already high to begin with. We tested this prediction by providing participants with a cue regarding their apparent level of conservativeness as judges.

In the control condition, we told participants after each mismatch judgment that most people who participated in a similar study also (correctly) thought that the item just presented was not the correct target. The alleged percentage of participants who presumably recognized the lures varied from 81% to 89%, because our previous results suggested that participants recognized the lures as incorrect about 80% of the time. Telling participants that most other participants were also correct should not be surprising to most of them, providing little or no diagnostic information over and above the baseline. Consequently, we predicted that the effect of seeing five versus nine lures on target misidentification would be pronounced in the control condition (see web appendix F for an illustration).

In the lax cue condition, we told participants that most people who participated in a similar study (incorrectly) thought that the item just presented was the correct target (we again varied the exact percentage from 81% to 89%). Given that most participants in our previous studies recognized the lures as incorrect most of the time, we expected this manipulation to lead participants to feel that they had already had a relatively stringent matching threshold, compared with most other participants. Consequently, we predicted that this manipulation would attenuate participants' tendency to increase their matching threshold further, thereby attenuating the effect of seeing five versus nine lures on target misidentification. A separate pretest (web appendix F) validated that this manipulation indeed led those in the lax cue condition to believe they were discerning more than those in the control condition ($M_{\text{lax cue}}$ = 5.67 vs. M_{control} = 4.98; $F(1, 74) = 4.56$, p = .036, Cohen's d = .49).

Finally, we measured the same focal dependent variable as in our previous studies: whether participants recognized the correct target when it was presented after the lures.

Results

A 2 (lures: five vs. nine) × 2 (social comparison cue: lax vs. control) linear probability model on target identification revealed a significant interaction ($t(316) = 2.03$, p = .043, Cohen's d = .23). In light of known issues with interpreting interaction effects in logistic regression models (Ai and Norton 2003, Winterich et al. 2013), we used a linear probability model to address the concerns (Angrist 2001; Angrist and Pischke 2008; Pischke 2011). The interaction effect derived from an adjusted logistic regression model (see Ai and Norton 2003; Hess, Hu, and Blair 2014) was also significant (Z = 2.06, p = .04). Replicating our findings, in the control condition (i.e., where people were told their identification accuracy was similar to that of most others), rejecting a greater number lures significantly reduced the identification accuracy for the correct target (nine lures: 61.4% vs. five lures: 76.5%; χ^2 (1) = 4.36, p = .037, Cohen's d = .24). In contrast, in the lax cue condition (i.e., where people were told their identification accuracy was already better than that of most others), the number of lure rejections did not influence recognition accuracy of the correct target (nine lures: 78.4% vs. five lures: 73.2%; χ^2 (1) = .57, p = .45). Figure 1 summarizes the results.

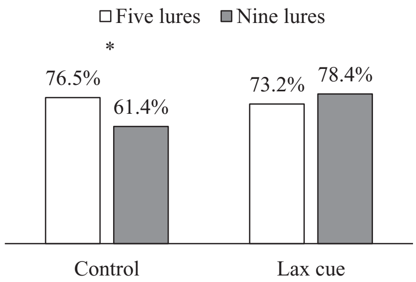
Note that, in line with our prediction, the effect of our social comparison manipulation was also significant when contrasting the two nine-lure conditions (lax cue condition: 78.4% vs. control condition: 61.4%; χ^2 (1) = 5.28, p = .022, Cohen's d = .26). When participants believed that their matching threshold was already relatively high to begin with, compared with most others, their tendency to further ratchet up their threshold was attenuated.

Discussion

Study 3 used a moderation of process approach to demonstrate the threshold escalation mechanism underlying the false rejection effect. Screening a longer sequence of lures increased false rejection of the correct target at baseline, or when nondiagnostic new information was provided (control cue condition), but not when new information implied

FIGURE 1

CORRECT IDENTIFICATION RATES OF TARGET IN STUDY 3



NOTE.—Asterisk (*) denotes a significant difference $p < .05$.

that one's internal matching threshold was already high to begin with (lax threshold cue condition).

STUDY 4: INDUCING THRESHOLD ESCALATION INCREASES FALSE REJECTION

Study 4 provides further evidence for the underlying mechanism using a different moderation-of-process manipulation. Whereas study 3 used a manipulation designed to attenuate threshold escalation, study 4 uses a manipulation designed to increase threshold escalation beyond its baseline level.

Specifically, for each lure in the lineup, participants received information about how similar that lure presumably was to the correct target, based on a purported objective algorithm. In the baseline threshold condition, participants were told that each of the lures they screened was similar to the correct target at a level comparable to participants' spontaneous judgments of similarity (based on a pretest). Because this does not provide new information to participants, we expected this condition to replicate our prior results.

In the high expected difficulty (i.e., stringent threshold) condition, each time participants saw a lure, they were told that the lure they had just seen was in fact very similar to the product they were looking for, heightening an overall sense of how similar the lures are to the target. This manipulation was designed to lead participants to expect the task to be more difficult, thereby encouraging them to adopt a stringent threshold (because only a stringent threshold can discern among very similar options; Benjamin and Bawa 2004). If threshold escalation is driving the false rejection of the correct target, then encouraging participants to adopt a high threshold (by leading them to expect that the task would be more difficult) should increase false rejection even when they did not screen many lures. Thus, we predicted false rejection to be equally high in this condition, regardless of the number of lures judged prior to the focal target. However, the manipulation of expected difficulty should have less of an effect in the many-lures condition, where people's threshold is already high as a result of screening more lures.

Method

Two hundred thirty-eight undergraduate students participated in exchange for course credit ($M_{\text{age}} = 21.0$, 52.1% women, no participants excluded from analyses). Participants were randomly assigned to one of the four following between-subjects conditions: 2 (lures: two vs. six) $\times$ 2 (expected difficulty: baseline vs. high). The procedure was similar to that of study 3, with two exceptions. First, we used robot vacuum cleaners instead of headphones as our product category. Participants saw either two or six robot vacuum cleaners in a random order before seeing the correct product (web appendix G).

Second, after mentally simulating a mismatch judgment for each lure in the sequence, participants received information on how similar that lure was to the correct target, as allegedly estimated by a Google algorithm.

In the baseline condition, each time participants saw a lure, they were told that based on a Google algorithm, the lure was somewhat similar to the correct target product. For example, participants were told that the "Google algorithm estimates that this product shares 57% of design features with the product you are looking for" after seeing each lure (expressed in terms of the percentage of shared features and ranging from 58% to 69%; web appendix G). We determined the alleged level of similarity for this condition based on a pretest ($N = 65$) which indicated that people perceived the lures to be approximately 60% similar to the correct target, on average, regardless of the length of the lineup ($M_{\text{six lures}} = 58.13\%$ vs. $M_{\text{two lures}} = 59.66\%$; $F(1, 63) = 0.11$, $p = .74$). Because this level of similarity is comparable to participants' spontaneous judgments of similarity, we expected our prior results to replicate in this condition.

In the high expected difficulty condition, after each time participants saw a lure and before they saw the next one, we told them that the item was very similar to the target in objective terms, based on a Google algorithm (the exact percentage of similarity ranged from 83% to 96%). We predicted that believing the lures were very similar to the target product would make participants perceive the identification task as more challenging overall, leading them to adopt a more stringent threshold. This, in turn, should strengthen the tendency to reject the correct target regardless of the number of lures judged. A separate pretest (web appendix G) validates that participants in the high expected difficulty condition expected that identifying the correct target would be significantly more difficult compared with those in the baseline condition ($M_{\text{high difficulty}} = 4.39$ vs. $M_{\text{baseline}} = 3.46$; $F(1, 77) = 7.89$, $p = .006$, Cohen's $d = .63$).

After making the identification decision on the correct target, participants indicated how similar they thought each lure was to the correct product, in terms of percentages, independent of the algorithm estimation as a manipulation check. We expected participants to perceive the lures as more similar to the target in the high expected difficulty condition, compared with the baseline condition, without a main effect of the number of lures on perceived similarity.

Results

A manipulation check revealed a main effect of the expected difficulty manipulation ($M_{\text{high difficulty}} = 72.84\%$ vs. $M_{\text{baseline}} = 53.69\%$, $F(1, 234) = 63.88$, $p < .001$, Cohen's $d = .46$), with no main effect or interaction involving the number of lures (main effect: $F(1, 234) = 0.1$, $p = .345$; interaction: $F(1, 234) = 0.01$, $p = .94$). We predicted that the expected difficulty would lead participants to adopt a stringent threshold, resulting in increased false rejection of the true target, regardless of the number of lures. A 2 (lures: two vs. six) × 2 (expected difficulty: baseline vs. high) linear probability model on target recognition revealed a marginally significant interaction effect ($t(234) = -1.68$, $p = .094$, Cohen's $d = .22$). An alternative logistic regression model revealed a similar result ($Z = 1.68$, $p = .092$).

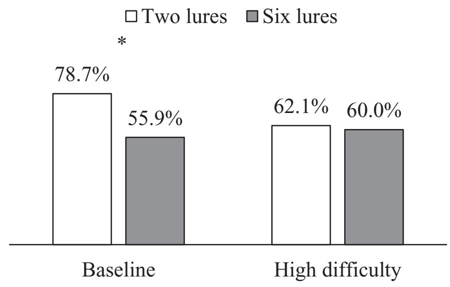
As predicted, the threshold escalation effect replicated in the baseline condition: seeing more lures before the focal target decreased the likelihood of correctly perceiving the target (six lures: 55.9% vs. two lures: 78.7%; χ^2 (1) = 7.08, $p = .008$, Cohen's $d = .30$). However, in the high expected difficulty condition, participants' correct identification rate was equally lower, regardless of the number of lures encountered (six lures: 60.0% vs. two lures: 62.1%; χ^2 (1) = .05, $p = .82$). See figure 2. The contrast between the high and baseline expected difficulty conditions within the two-lures level was also significant (high: 62.1% vs. baseline: 78.7%; χ^2 (1) = 3.96, $p = .047$, Cohen's $d = .22$), supporting our prediction that inducing participants to adopt a more stringent threshold increased false rejection rates.

Discussion

Study 4 supports our threshold escalation process hypothesis by demonstrating that encouraging participants to adopt a stringent threshold, by leading them to expect that it would be difficult to discern the target, increased false rejection when the sequence of lures was short (i.e., when participants' threshold was low at baseline). The same manipulation had no incremental effect when the sequence of

FIGURE 2

CORRECT IDENTIFICATION RATES OF TARGET IN STUDY 4



NOTE.—Asterisk (*) denotes a significant difference $p < .05$.

lures was long (i.e., when the matching threshold was already higher as a result of our proposed process).

Taken together, studies 3 and 4 provide convergent evidence that threshold placement is driving the effect, by showing that experimentally lowering the threshold reduces false rejection rates for those in the more-lures condition (study 3) and that exogenously increasing the threshold increases false rejection for those in the fewer-lures condition (study 4).

The results from both studies cast doubt on the possibility that experimental demand, simple error, or memory decay (perceptual interference) might be driving the effect, since the actual lures participants saw were held constant, but only the supplementary information changed (i.e., social comparison cue, study 3; expected difficulty, study 4). None of these alternative accounts explains the observed interactions. Using different products further bolsters the generalizability.

STUDY 5: BOUNDARY CONDITION

Study 5 examines a boundary condition for the threshold escalation effect. Our conceptualization is based on the premise that people assess the target's degree of match based on a continuous notion of similarity or familiarity. Namely, the focal stimulus can feel more or less right, and whether participants perceive it as the correct target depends on whether this perceived level of familiarity exceeds their internal matching threshold. It follows that identification accuracy should be less sensitive to shifts in the placing of the matching threshold when the target includes a distinctive feature, one that can be identified in a more binary sense (i.e., the target either has it or does not). The subjective perception of *relative* similarity or familiarity should be less pertinent when the target can be identified based on an objective, unique cue not possessed by the lures. In such cases, people are likely to rely on a distinctiveness heuristic (Dodson and Schacter 2002;

Schacter, Israel, and Racine 1999), where they identify an item based on the absence or presence of an expected distinctive cue.

Study 5 tests this prediction by including a new condition, in which we added a subtle but distinctive cue to the focal target. Study 5 also used a different experimental paradigm (i.e., website search paradigm) to bolster the generalizability and robustness. For brevity, we describe the specific details of the study and all the ancillary analyses in web appendix H.

Method

Three hundred seventy-one participants from MTurk were randomly assigned to a condition in a 2 (lures: two vs. six) $\times$ 2 (distinctive cue: present vs. absent) between-subjects design. Those in the cue-present (absent) condition saw a target product that had (did not have) an unobtrusive yet noticeable logo on one side.

To bolster generalizability, we used a modified procedure for the lineup that resembled an online search. Participants saw the same pair of headphones used in previous studies and were asked to assume that they were searching for the product online. All participants then saw eight hyperlinks, presumably reflecting their search results, each linked to a pop-up window with a picture of a product. We asked participants to go through the options in order. The third (seventh) link in the two-lures condition (six-lures condition) contained the correct target product, while the other links showed a lure product (see web appendix H for details). Participants then made spontaneous (i.e., not simulated) correct/incorrect judgments for each option presented in the lineup. We predicted that the threshold escalation effect would not hold when the target has a distinctive feature.

Results

The results reveal a two-way interaction ($t(367) = 2.40$, $p = .017$, Cohen's $d = .25$). Replicating our prior findings, the threshold escalation effect emerged in the cue-absent condition: participants were less likely to accurately identify the true target in the six-lure (58.5%) than in the two-lure condition (83.8%; χ^2 (1) = 14.83, $p < .001$, Cohen's $d = .41$). Consistent with our boundary condition prediction, this effect was attenuated in the cue-present condition (six lures: 78.4% vs. two lures: 82.9%; $\chi^2(1) = .58$, $p = .45$). We ran several supplementary analyses to provide additional support for our findings and to rule out alternative accounts including mere-inaccuracy tests as in study 1, endogeneity tests using a matched-choice paradigm (Gal and Liu, 2011), and calculation of discrimination sensitivity index (d') and a threshold placement index (C). We elaborate on these analyses in web appendix H.

Discussion

Study 5 replicates our main finding using a different design. Further, the study shows a boundary condition: when participants can identify the target based on a distinctive feature, without relying on subjective similarity, the effect of sequence length is attenuated.

In a follow-up study (web appendix I), we tested how the actual degree of similarity between the lures and the target influenced our effect. Conceptually consistent with study 5, making the lures less similar to the target decreased the overall rate of false rejections (i.e., failures to correctly identify the target). Please refer to the web appendix for a detailed discussion of the results.

GENERAL DISCUSSION

Consumers often try to visually identify a specific, previously encountered product among multiple lookalikes, an experience we label a product lineup. Extending prior research on visual recognition, we show that the more participants sequentially screen lookalike options (i.e., lures), the more conservative but—ironically—less accurate judges they become. Consequently, they are more likely to fail to recognize the product for which they have been looking when it finally appears in the lineup.

We proposed that this happens because each time consumers evaluate an item in a product lineup and determine that it is not the option for which they have been looking, they draw an implicit inference that the correct target should feel even more familiar or right. This causes their internal matching threshold (i.e., the subjective degree of familiarity or rightness they expect to experience with the true target) to gradually escalate the more they have evaluated and rejected similar but incorrect lures. Consequently, the longer they search and encounter more lures, the more conservative judges they become and the more likely they are to fail to recognize the true target when it reappears in the lineup.

Five main studies and four follow-up studies (see web appendix J for a summary), reported in this article and the web appendix, support the proposed effect and rule out alternative explanations based on mere error, memory decay, overload, doubt, confidence, and rejection inertia. The studies also cast doubt on the possibility that participants strategically adjusted their threshold in anticipation of the number of lures, by either not mentioning the length of product lineup to participants (e.g., study 1) or holding the total number of items in the product lineup constant across conditions (study 5).

Our findings suggest that threshold escalation reflects the sequential nature of the identification process. One may wonder whether it persists when the options are presented simultaneously, as is often the case when search results are displayed on a single web page. We believe that the findings apply to many such situations. First, consumers often examine images sequentially, even when several initial results are displayed on a single screen, when they click on each image in turn to enlarge it or open multiple browser tabs. Second, viewing multiple items on a single screen is inconvenient when people use smartphones, thereby leading to sequential evaluation. Third, even when consumers view multiple large-enough images on a single display, they ultimately attend to each image separately in turn (Stewart, Chater, and Brown 2006). Thus, the consequences of sequential evaluation are likely to persist across many different everyday scenarios where consumers rely on their memory to visually find a specific option. That said, to the extent that consumers are engaging in a truly simultaneous evaluations (e.g., when comparing options in a multi-attribute matrix), threshold escalation may not apply. Such simultaneous comparisons tend to be more analytical and therefore characteristic of later stages in product search (e.g., when consumers have narrowed their search to a few leading alternatives and are deliberating on pros and cons). They are less characteristic of consumers' initial attempts to identify previously seen options based on visual memory.

Theoretical and Practical Implications

Our research contributes to the consumer search literature by defining the notion of product lineups and examining factors that influence consumers' recognition accuracy. In addition, our research extends existing theories of visual recognition by examining a novel mechanism through which perceivers' internal matching threshold may escalate in sequential recognition tasks. To our best knowledge, our research is among the first to explore how prior mismatch judgments impact downstream identification accuracy.

The current research joins a growing body of literature that examines judgment bubbles, where perceptual or evaluative judgment biases are exacerbated the more consumers persist or invest effort in a task. Prior research has shown, for example, how trivial decisions may become more difficult the harder one tries to solve them (Sela and Berger 2012), and how estimation bubbles may lead to absurd results (Simmons 2018). Our findings demonstrate another context where the longer people try to form a particular judgment, the less likely they are to generate an accurate response, thereby echoing the notion of judgment bubbles.

More broadly, our findings contribute to the literature on eyewitness identification in police lineups. Research has identified several biases in suspect identification (Reisberg 2014). For example, eyewitnesses tend to use a more stringent matching threshold when potential culprits are presented sequentially compared to when they are presented simultaneously (Meissner et al. 2005). The current findings suggest that the number of items dismissed before the true target, in addition to their presentation format, can also elevate threshold placement.

From a practical perspective, our findings underscore the importance of matching items to appear at the top of search engine results when consumers are looking for a specific option. This has implications for online ad bidding and other product search algorithms, because threshold escalation makes bidding higher for an advantageous page placement even more critical. The findings also suggest that sellers may benefit from including a rich array of keywords pertaining to the product's visual cues, thereby allowing searching consumers to find a better visual match more easily. Marketers may also benefit from drawing consumers' attention to visually distinct logos and cues, to attenuate the negative effects of threshold escalation and minimize the likelihood that consumers miss the product.

For consumers, the results suggest that reviewing extensive product lineups in search of a specific product may, ironically, reduce their likelihood of success. One potential strategy for mitigating this experience may be to avoid prolonged search sessions and to take a step back from the task. Short breaks or momentary task switching can change information processing modes (Liu 2008; Sela and Berger 2012), which may also help reset the internal threshold.

Future Research

The current article focuses on recognition, holding preference constant, but one interesting direction for future research is that the same threshold escalation mechanism may also influence preference construction. Encountering a sequence of insufficiently satisfying options (i.e., below the subjective acceptance threshold) may lead people to adopt an increasingly high threshold due to a similar inference, where they expect an acceptable option to provide a higher level of satisfaction with every option they reject. For example, zapping through a longer sequence of TV or radio channels may lead consumers to become pickier, and consequently less likely to eventually find a channel they feel they like. Future research may examine whether our process may account for such a pattern.

One may also wonder about the consequences that follow the recognition decision. Might adopting a higher matching threshold, for example, due to screening more lures, increase willingness to pay or choice satisfaction? How does failing to visually identify options in a product lineup influence consumers' reliance on other decision rules, such as opting for safer, more expensive, or more established brands (Simonson 1992)? Future research may examine these and other important questions pertinent to product lineups.

DATA COLLECTION INFORMATION

The first author collected and analyzed the data under the supervision of the second author. Data for all studies was collected between spring 2017 and fall 2018, with the exception of the study 5 follow-up, which was collected during spring 2019. Data for studies 3 and 5 was collected on MTurk. Participants for study 1 were recruited on the University of Florida campus. Study 2 and its second follow-up was run by research assistants at the University of Florida Behavior Lab. Study 4 was conducted at the Virginia Tech behavioral lab, administered by a lab manager who was supervised by the first author. Data for all other follow-up studies were collected on MTurk. A subset of participants for the pilot study were collected through social media (see web appendix A).

APPENDIX A

Wireless computer mouse devices in study 1

Two wireless computer mouse devices provided to participants (i.e., correct target):

Lure wireless computer mouse devices (two or six presented, in random order, depending on the condition):

REFERENCES

Ai, Chunrong and Edward C. Norton (2003), "Interaction Terms in Logit and Probit Models," *Economics Letters*, 80 (1), 123–9.

Angrist, Joshua D. (2001), "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business & Economic Statistics*, 19 (1), 2–28.

Angrist, Joshua D. and Jörn-Steffen Pischke (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, NJ: Princeton University Press.

Ashby, F. Gregory and Nancy A. Perrin (1988), "Toward a Unified Theory of Similarity and Recognition," *Psychological Review*, 95 (1), 124–50.

Beckwith, Neil E. and Donald R. Lehmann (1975), "The Importance of Halo Effects in Multi-Attribute Attitude Models," *Journal of Marketing Research*, 12 (3), 265–75.

Benjamin, Aaron S. and Sameer Bawa (2004), "Distractor Plausibility and Criterion Placement in Recognition," *Journal of Memory and Language*, 51 (2), 159–72.

Bettman, James R. (1979), "Memory Factors in Consumer Choice: A Review," *Journal of Marketing*, 43 (2), 37–53.

Brown, J., V. J. Lewis, and A. F. Monk (1977), "Memorability, Word Frequency and Negative Recognition," *Quarterly Journal of Experimental Psychology*, 29 (3), 461–73.

Bullock, John G., Donald P. Green, and Shang E. Ha (2010), "Yes, but What's the Mechanism? (Don't Expect an Easy Answer)," *Journal of Personality and Social Psychology*, 98 (4), 550–8.

Cesario, Joseph, Heidi Grant, and E. Tory Higgins (2004), "Regulatory Fit and Persuasion: Transfer from 'Feeling Right,'" *Journal of Personality and Social Psychology*, 86 (3), 388–404.

Diehl, Kristin (2005), "When Two Rights Make a Wrong: Searching Too Much in Ordered Environments," *Journal of Marketing Research*, 42 (3), 313–22.

Diehl, Kristin and Cait Poynor (2010), "Great Expectations?! Assortment Size, Expectations, and Satisfaction," *Journal of Marketing Research*, 47 (2), 312–22.

Dodson, Chad S. and Daniel L. Schacter (2002), "When False Recognition Meets Metacognition: The Distinctiveness Heuristic," *Journal of Memory and Language*, 46 (4), 782–803.

Dolese, Melissa, Debra Zellner, Marsha Vasserman, and Scott Parker (2005), "Categorization Affects Hedonic Contrast in the Visual Arts," *Bulletin of Psychology and the Arts*, 5 (2), 21–5.

Downing, James W., Charles M. Judd, and Markus Brauer (1992), "Effects of Repeated Expressions on Attitude Extremity," *Journal of Personality and Social Psychology*, 63 (1), 17–29.

Eich, Janet M. (1982), "A Composite Holographic Associative Recall Model," *Psychological Review*, 89 (6), 627–61.

Estes, William Kaye and W. Todd Maddox (1995), "Interactions of Stimulus Attributes, Base Rates, and Feedback in Recognition," *Journal of Experimental Psychology-Learning Memory and Cognition*, 21 (5), 1075–94.

Fazio, Russell H., Jeaw-Mei Chen, Elizabeth C. McDonel, and Steven J. Sherman (1982), "Attitude Accessibility, Attitude-Behavior Consistency, and the Strength of the Object-Evaluation Association," *Journal of Experimental Social Psychology*, 18 (4), 339–57.

Ford, Thomas E. and Erik P. Thompson (2000), "Preconscious and Postconscious Processes Underlying Construct Accessibility Effects: An Extended Search Model," *Personality and Social Psychology Review*, 4 (4), 317–36.

Gal, David, and, Wendy. Liu (2011), "Grapes of Wrath: The Angry Effects of Self-Control," Journal of Consumer Research, 38 (3), 445–58.

Gillund, Gary and Richard M. Shiffrin (1984), "A Retrieval Model for Both Recognition and Recall," *Psychological Review*, 91 (1), 1–67.

Green, David M. and John A. Swets (1966), *Signal Detection Theory and Psychophysics*, New York: Wiley.

Hauser, John R., Songting Dong, and Min Ding (2014), "Self-Reflection and Articulated Consumer Preferences," *Journal of Product Innovation Management*, 31 (1), 17.

Hess, James D., Ye Hu, and Ed Blair (2014), "On Testing Moderation Effects in Experiments Using Logistic Regression," working paper, University of Houston, Bauer College of Business, Houston, TX 77004.

Hirshman, Elliot (1995), "Decision Processes in Recognition Memory: Criterion Shifts and the List-Strength Paradigm," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21 (2), 302–13.

Hoch, Stephen J., Eric T. Bradlow, and Brian Wansink (1999), "The Variety of an Assortment," *Marketing Science*, 18 (4), 527–46.

Hockley, William E. (2008), "The Picture Superiority Effect in Associative Recognition," *Memory & Cognition*, 36 (7), 1351–9.

Holland, Rob W., Bas Verplanken, and Ad van Knippenberg (2003), "From Repetition to Conviction: Attitude Accessibility as a Determinant of Attitude Certainty," *Journal of Experimental Social Psychology*, 39 (6), 594–601.

Huber, Joel and Morris B. Holbrook (1979), "Using Attribute Ratings for Product Positioning: Some Distinctions among Compositional Approaches," *Journal of Marketing Research*, 16 (4), 507–16.

Judd, Charles M. and Markus Brauer (1995), "Repetition and Evaluative Extremity," in *Ohio State University Series on Attitudes and Persuasion, Vol. 4*, Attitude Strength: Antecedents and Consequences, ed. Richard E. Petty and Jon A. Krosnick, Hillsdale, NJ: Erlbaum, 43–72.

Krueger, Lester E. (1978), "A Theory of Perceptual Matching," *Psychological Review*, 85 (4), 278–304.

Liu, Wendy (2008), "Focusing on Desirability: The Effect of Decision Interruption and Suspension on Preferences," *Journal of Consumer Research*, 35 (4), 640–52.

Megreya, Ahmed M. and A. Mike Burton (2007), "Hits and False Positives in Face Matching: A Familiarity-based Dissociation," *Perception & Psychophysics*, 69 (7), 1175–84.

Meissner, Christian A., Colin G. Tredoux, Janat F. Parker, and Otto H. MacLin (2005), "Eyewitness Decisions in Simultaneous and Sequential Lineups: A Dual-Process Signal Detection Theory Analysis," *Memory & Cognition*, 33 (5), 783–92.

Miller, Michael B. and George L. Wolford (1999), "Theoretical Commentary: The Role of Criterion Shift in False Memory," *Psychological Review*, 106 (2), 398–405.

Moorthy, Sridhar, Brian T. Ratchford, and Debabrata Talukdar (1997), "Consumer Information Search Revisited: Theory and Empirical Analysis," *Journal of Consumer Research*, 23 (4), 263–77.

Mussweiler, Thomas (2003), "Comparison Processes in Social Judgment: Mechanisms and Consequences," *Psychological Review*, 110 (3), 472–89.

Norris, Claire E. and Andrew M. Colman (1992), "Context Effects on Recall and Recognition of Magazine Advertisements," *Journal of Advertising*, 21 (3), 37–46.

Nosofsky, Robert M. and Safa R. Zaki (2003), "A Hybrid-Similarity Exemplar Model for Predicting Distinctiveness Effects in Perceptual Old-New Recognition," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29 (6), 1194–209.

Novemsky, Nathan, Ravi Dhar, Norbert Schwarz, and Itamar Simonson (2007), "Preference Fluency in Choice," *Journal of Marketing Research*, 44 (3), 347–56.

Oppenheimer, Daniel M., Tom Meyvis, and Nicolas Davidenko (2009), "Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power," *Journal of Experimental Social Psychology*, 45 (4), 867–72.

Palmer, Matthew A. and Neil Brewer (2012), "Sequential Lineup Presentation Promotes Less-Biased Criterion Setting but Does Not Improve Discriminability," *Law and Human Behavior*, 36 (3), 247–55.

Parker, Scott, Jesse Bascom, Brian Rabinovitz, and Debra Zellner (2008), "Positive and Negative Hedonic Contrast with Musical Stimuli," *Psychology of Aesthetics, Creativity, and the Arts*, 2 (3), 171–4.

Pischke, Jörn-Steffen (2011), "Money and Happiness: Evidence from the Industry Wage Structure," NBER Working Paper No. 17056.

Pleskac, Timothy J. (2007), "A Signal Detection Analysis of the Recognition Heuristic," *Psychonomic Bulletin & Review*, 14 (3), 379–91.

Reisberg, Daniel (2014), *The Science of Perception and Memory: A Pragmatic Guide for the Justice System*, New York: Oxford University Press.

Rhodes, Matthew G. and Larry L. Jacoby (2007), "On the Dynamic Nature of Response Criterion in Recognition Memory: Effects of Base Rate, Awareness, and Feedback," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33 (2), 305–20.

Rothschild, Michael (1978), "Searching for the Lowest Price When the Distribution of Prices Is Unknown," in *Uncertainty in Economics*, ed. Peter Diamond and Michael Rothschild, New York: Academic Press, 425–54.

Rowley, Jennifer (2000), "Product Search in E-Shopping: A Review and Research Propositions," *Journal of Consumer Marketing*, 17 (1), 20–35.

Schacter, Daniel L. (1999), "The Seven Sins of Memory: Insights from Psychology and Cognitive Neuroscience," *American Psychologist*, 54 (3), 182–203.

Schacter, Daniel L., Lana Israel, and Carrie Racine (1999), "Suppressing False Recognition in Younger and Older Adults: The Distinctiveness Heuristic," *Journal of Memory and Language*, 40 (1), 1–24.

Schrift, Rom Y., Oded Netzer, and Ran Kivetz (2011), "Complicating Choice," *Journal of Marketing Research*, 48 (2), 308–26.

Schwarz, Norbert (2006), "Feelings, Fit, and Funny Effects: A Situated Cognition Perspective," *Journal of Marketing Research*, 43 (1), 20–3.

Segal, Sydney J. and Vincent Fusella (1970), "Influence of Imaged Pictures and Sounds on Detection of Visual and Auditory Signals," *Journal of Experimental Psychology*, 83 (3, Pt.1), 458–64.

Sela, Aner and Jonah Berger (2012), "Decision Quicksand: How Trivial Choices Suck Us In," *Journal of Consumer Research*, 39 (2), 360–70.

Sela, Aner and Baba Shiv (2009), "Unraveling Priming: When Does the Same Prime Activate a Goal versus a Trait?" *Journal of Consumer Research*, 36 (3), 418–33.

Sherif, Muzafer and Carl I. Hovland (1961), *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*, New Haven, CT: Yale University Press.

Shoots-Reinhard, Brittany L., Derek D. Rucker, Richard E. Petty, and Richard Shakarchi (2014), "Not All Contrast Effects Are Created Equal: Extent of Processing Affects Contrast Strength," *Journal of Applied Social Psychology*, 44 (8), 523–35.

Simmons, Joseph P. (2018), "Elephants Weigh More Than…Elephants: How Prediction Momentum Arises from Sequential and Independent Judgments," working paper, Operations, Information, and Decisions Department, Wharton School, University of Pennsylvania, Philadelphia, PA 19104

Simonson, Itamar (1992), "The Influence of Anticipating Regret and Responsibility on Purchase Decisions," *Journal of Consumer Research*, 19 (1), 105–18.

Singer, Murray, Nathalie Gagnon, and Eric Richards (2002), "Strategies of Text Retrieval: A Criterion Shift Account," *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 56 (1), 41–57.

Singer, Murray and John T. Wixted (2006), "Effect of Delay on Recognition Decisions: Evidence for a Criterion Shift," *Memory & Cognition*, 34 (1), 125–37.

Snodgrass, Joan Gay, Rochelle Volvovitz, and Eleanor Radin Walfish (1972), "Recognition Memory for Words, Pictures, and Words + Pictures," *Psychonomic Science*, 27 (6), 345–7.

Spencer, Steven J., Mark P. Zanna, and Geoffrey T. Fong (2005), "Establishing a Causal Chain: Why Experiments Are Often More Effective than Mediational Analyses in Examining Psychological Processes," *Journal of Personality and Social Psychology*, 89 (6), 845–51.

Stewart, Neil, Nick Chater, and Gordon D. A. Brown (2006), "Decision by Sampling," *Cognitive Psychology*, 53 (1), 1–26.

Stigler, George J. (1961), "The Economics of Information," *Journal of Political Economy*, 69 (3), 213–25.

Stretch, Vincent and John T. Wixted (1998), "On the Difference between Strength-based and Frequency-based Mirror Effects in Recognition Memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24 (6), 1379–96.

Tesser, Abraham and Christopher Leone (1977), "Cognitive Schemas and Thought as Determinants of Attitude Change," *Journal of Experimental Social Psychology*, 13 (4), 340–56.

Tesser, Abraham, Leonard Martin, and Marilyn Mendolia (1995), "The Impact of Thought on Attitude Extremity and Attitude-Behavior Consistency," in *Ohio State University Series on Attitudes and Persuasion, Vol. 4, Attitude Strength: Antecedents and Consequences*, ed. Richard E. Petty and Jon A. Krosnick, Hillsdale, NJ: Erlbaum, 73–92.

Titus, Philip A. and Peter B. Everett (1995), "The Consumer Retail Search Process: A Conceptual Model and Research Agenda," *Journal of the Academy of Marketing Science*, 23 (2), 106–19.

Tykocinski, Orit E. and Thane S. Pittman (1998), "The Consequences of Doing Nothing: Inaction Inertia as Avoidance of Anticipated Counterfactual Regret," *Journal of Personality and Social Psychology*, 75 (3), 607–16.

Verde, Michael F. and Caren M. Rotello (2007), "Memory Strength and the Decision Process in Recognition Memory," *Memory & Cognition*, 35 (2), 254–62.

Wänke, Michaela, Gerd Bohner, and Andreas Jurkowitsch (1997), "There Are Many Reasons to Drive a BMW: Does Imagined Ease of Argument Generation Influence Attitudes?" *Journal of Consumer Research*, 24 (2), 170–7.

Warren, Richard M. (1985), "Criterion Shift Rule and Perceptual Homeostasis," *Psychological Review*, 92 (4), 574–84.

Whittlesea, Bruce W. A., Larry L. Jacoby, and Krista Girard (1990), "Illusions of Immediate Memory: Evidence of an Attributional Basis for Feelings of Familiarity and Perceptual Quality," *Journal of Memory and Language*, 29 (6), 716–32.

Winkielman, Piotr and Norbert Schwarz (2001), "How Pleasant Was Your Childhood? Beliefs about Memory Shape Inferences from Experienced Difficulty of Recall," *Psychological Science*, 12 (2), 176–9.

Winterich, Karen Page, Karl Aquino, Vikas Mittal, and Richard Swartz (2013), "When Moral Identity Symbolization Motivates Prosocial Behavior: The Role of Recognition and Moral Identity Internalization," *Journal of Applied Psychology*, 98 (5), 759–70.

Wixted, John T. (2007), "Dual-Process Theory and Signal-Detection Theory of Recognition Memory," *Psychological Review*, 114 (1), 152–76.

Xu, Alison Jing and, Wyer Robert S. Jr. (2012), "The Role of Bolstering and Counterarguing Mind-Sets in Persuasion," *Journal of Consumer Research*, 38 (5), 920–32.

Zellner, Debra A., Elizabeth A. Rohm, Terri L. Bassetti, and Scott Parker (2003), "Compared to What? Effects of Categorization on Hedonic Contrast," *Psychonomic Bulletin & Review*, 10 (2), 468–73.

Zwick, Rami, Amnon Rapoport, Alison King Chung Lo, and A. V. Muthukrishnan (2003), "Consumer Sequential Search: Not Enough or Too Much?" *Marketing Science*, 22 (4), 503–19.