

# A Preference for Revision Absent Objective Improvement

Ximena Garcia-Rada  
Leslie K. John  
Ed O'Brien  
Michael I. Norton

Working Paper 19-087



# A Preference for Revision Absent Objective Improvement

Ximena Garcia-Rada  
Harvard Business School

Leslie K. John  
Harvard Business School

Ed O'Brien  
University of Chicago

Michael I. Norton  
Harvard Business School

**Working Paper 19-087**

Copyright © 2019, 2020 by Ximena Garcia-Rada, Leslie K. John, Ed O'Brien, and Michael I. Norton.

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

Funding for this research was provided in part by Harvard Business School.

A Preference for Revision Absent Objective Improvement

XIMENA GARCIA-RADA

LESLIE K. JOHN

ED O'BRIEN

MICHAEL I. NORTON

Ximena Garcia-Rada ([xgarciarada@hbs.edu](mailto:xgarciarada@hbs.edu)) is a doctoral candidate in marketing at the Harvard Business School, Harvard University; Leslie K. John ([ljohn@hbs.edu](mailto:ljohn@hbs.edu)) is a Marvin Bower Associate Professor of Business Administration at the Harvard Business School, Harvard University; Ed O'Brien ([eob@chicagobooth.edu](mailto:eob@chicagobooth.edu)) is an Associate Professor of Behavioral Science at the University of Chicago Booth School of Business; Michael I. Norton ([mnorton@hbs.edu](mailto:mnorton@hbs.edu)) is the Harold M. Brierly Professor of Marketing at the Harvard Business School, Harvard University. The authors thank Derek Koehler for helpful feedback and Holly Howe, Shannon Sciarappa, Trevor Spelman, and Shun Wang for research assistance. To collect data for study 1A, we worked closely with the team in a professional development office at a business school and are grateful for their support. Please address correspondence to Ximena Garcia-Rada.

## CONTRIBUTION STATEMENT

Companies regularly release revised products (e.g., new editions of books, technological updates), and consumers often purchase those revised products—presumably because revising things makes them better. But might consumers prefer revised products even when companies create products that do not represent improvement over their predecessors? If so, what company actions might increase or attenuate consumers' trust that revisions signal improvement? The current investigation demonstrates that consumers are influenced by mere revision status (holding constant chronological newness): Consumers deem revised products as superior to unrevised counterparts, even in cases where objective improvement is absent, and even when flaws are evident. We further show that this preference is driven by assumptions that firms act in good faith when releasing revisions, a preference that diminishes when those assumptions of good faith are called into question.

This research has important marketing implications because consumers live in a world of ever-evolving offerings. Products that are objectively unchanged (or in some cases, even made worse) in the revision process may nonetheless be adopted, provided consumers believe they are experiencing a “revised” version. Conversely, products that are objectively improved in the revision process may go unappreciated, if the fact that they have been revised is not made clear to consumers.

This research contributes to several literatures in consumer behavior and decision-making. First, this research extends work on the links between expectations and experience, demonstrating how expectations change consumers' perceptions and experiences when they learn that products have been revised: consumers literally fail to notice their flaws. Second, as we expand upon in the General Discussion, this over-generalized preference for revision joins an

emerging collection of related phenomena that share two features: a sequencing and a framing component. Our work suggests that trust in the creator's intentions may be a critical feature that determines when consumers will prefer different temporal sequences (i.e., when consumers will prefer originals versus revisions).

## ABSTRACT

Things change. Things also *get* changed—often. Why? The obvious reason is that revising things often makes them better. We document a less obvious reason: revising things makes consumers *think* they are better, even absent objective improvement. Eleven studies document the preference for revision and provide insight into its psychological underpinnings. Studies 1A-1C document the effect among MBA students engaged in a resume revision process, while Studies 2 and 3 demonstrate the effect both experientially (eating candy) and via choice (of pens). Revisions are preferred even when trivially different from their predecessors (study 4A), and even when there is *no* difference between original and revised versions (study 4B). Consumers overgeneralize their belief that revisers intend to improve their creations, and that revisions represent the successful fruition of those intentions (studies 5A and 5B); as a result, consumers are relatively uncritical of flaws in purportedly revised products (a video game; study 6), unless cued to doubt that revisers can be trusted (study 7).

*Keywords:* revision, change, sequences, inferences.

Things change. Things also *get* changed—often. For example, companies regularly release revised products, editions of books, director’s cuts of movies, and technological updates. Individuals regularly tweak recipes, edit resumes, and revise manuscripts, sometimes endlessly. Why? The obvious reason is that revising things often makes them better. In the current research, we document a less obvious reason: Revising things makes people *think* they are better, absent objective improvement. In a public demonstration providing anecdotal support for this idea (see TIME 2017), city passersby were invited for a sneak-peak to play with the soon-to-be-released iPhone X (to be released later that month, in November 2017). Unsurprisingly, many raved about the phone’s sleek new feel and performance; in reality, all had been handed a well-worn iPhone 4 (released all the way back in June 2010). Even when people have full information—when they can check for themselves by seeing and using unimproved revisions while forming judgments—they can exhibit a preference for revision.

Of course, it is often logical to assume that revisions are better than originals: companies often revise their products to improve their quality. In many cases, however, changes can be merely cosmetic or superficial. If the goal of a revision is to simply draw renewed attention to a product, a firm may simply change its coloring or packaging. To be sure, the difference may be substantive in quality: the goal of a revision may be to improve the product, and the revision may reflect the fruition of this goal. But, a quality difference could also go in the other direction. For example, “planned obsolescence” is the practice of introducing products with short useful lives to spur faster repurchase (Bulow 1986), and “product versioning” is the strategy of deliberately degrading features in existing models to reduce the functionality of a product (Gershoff, Kivetz, and Keinan 2012). In the current research, we show that even in such cases when no objective improvements are made, consumers exhibit a preference for revision, stemming from an

overgeneralization of an often-reasonable inference (Baron 1990): that revisers are in good faith intending to improve their creations and that the revision realizes those intentions. As a result, consumers place undue faith in revisions—in a sense, forgoing critical assessment—leading to the perception that revisions are superior to their originals.

## **THEORETICAL BACKGROUND**

Our theoretical account of why consumers might prefer revisions absent objective improvement begins with the fact that human judgment is inherently comparative (Hsee, Loewenstein, Blount, and Bazerman 1999; Moore and Small 2007), and that encountering a revised version prompts a natural comparison group: the original version. But comparing quality can be difficult, especially when encountering stimuli for the first time. Under such uncertainty, consumers seek out other routes to simplify the task—namely, by prioritizing whatever attributes are most easily evaluable (Hsee, 1996; Hsee and Zhang 2010) and by drawing on peripheral contextual cues as opposed to more central content (Petty and Cacioppo 1986). Thus, at the broadest level, we suggest revision labels likely serve as an easily evaluable, peripheral cue to aid one’s judgment. Consumers may incorporate the mere fact that a product has been “revised” into their assessments of its quality. This possibility is consistent with other research that documents the effects of expectations on consumption experiences of identical stimuli (e.g., drinking a nasty “vinegar” beer makes it taste worse; drinking a “pricey” wine makes it taste better; eating a “fatty” dessert makes it taste more decadent: see Deliza and MacFie 1996; Hoch and Ha 1986; Lee, Frederick, and Ariely 2006).

Going beyond this broader reasoning, however, more specific theorizing is needed to account for our hypothesis that revision labels may have a distinctly *positive* effect on perceived

quality, as well as the possibility that they wield influence even when consumers should have an easier time with the task (i.e., when it should be obvious that a revision label is merely cosmetic and that no objective improvements have been made). In daily life, incidental preferences that are unrelated to the “revision” aspect per se may contribute to this effect (e.g., revised products are often newer products, and consumers may prefer newer versions for purposes of variety or status signaling: Henard and Szymanski 2001; Ratner and Kahn 2002)—but such preferences also cannot account for our hypothesis, as we theorize (and test for) a positive effect of revision labels while holding constant the chronological newness of the product in question.

First, our more specific theorizing draws on research documenting that people place high value on improvement often for mere sake of change. For example, people prefer sequences of improving outcomes (e.g., raises in annual income) over unchanging but higher quality outcomes (e.g., a flat but larger income, Klein and O’Brien 2017; Loewenstein and Prelec 1993); people actively denigrate their earlier selves so as to seem like they are improving over time (Wilson and Ross 2001); and people express more interest in others who have high potential to improve than in others who have already made the same achievements (Tormala, Jia, and Norton 2012). Moreover, O’Brien and Kardas (2016) documented a semantic association between the concept of change in oneself and distinctly *positive* reactions: participants who were asked to reflect on how they had changed as people over time, provided with no specific definition of “change,” brought to mind only ways they had *improved*—despite easily bringing to mind decline when prompted. Thus, although change can take many forms (including a product’s decline, as in planned obsolescence and product versioning), we propose that revision labels generally prompt consumers to approach a “revised” stimulus with positive expectations in mind.

Second, we posit that consumers' positive perceptions of revision are undergirded by a lay belief that revisers generally intend to improve their creations. We ground this prediction in previous research suggesting that people have general orienting systems: sets of global beliefs that guide their perceptions and interpretations (Antonovsky 1987; Epstein 1991; Janoff-Bulman and Frantz 1997; Mischel and Morf 2003; Park 2010). People commonly believe that the world is just (Lerner 1980), benevolent (Taylor and Brown 1988, 1994; Wortman and Silver 1992), predictable, coherent, and controllable (Janoff-Bulman and Frieze 1983; Wortman 1983). Applied to revision, we posit that these orienting systems lead consumers to assume that revisers have benevolent intentions (all else equal)—that is, to sincerely intend to improve their creations—and in turn, trust that revisers successfully executed those intentions. People often conflate others' intentions to engage in behavior with their implementation of that behavior (Fishbein 1980; Fishbein and Ajzen 1976); when making predictions about future behavior, for example, people perceive intentions to be strongly predictive of behavior, when in reality, they are typically only weakly predictive (Koehler, White, and John 2011). As a result, we suggest that an (overgeneralized) lay belief in good intentions also translates into an (overgeneralized) belief in good outcomes: that creators' intentions to improve products actually results in improvement.

This line of theorizing suggests a novel attentional mechanism underlying our hypothesis. If, on average, revision labels prompt positive expectations of change, such that consumers assume (i) revisers must have had good intentions when revising, and (ii) their intentions to improve must have translated into successful improvement, then it follows that consumers may approach revisions with more trust and less scrutiny than they would normally express merely due to the revision label. Consumers may assume that certain flaws have been fixed, and thus, even absent objective improvement, they may be less attentive to the presence of those flaws

while testing out a “revised” product—thereby concluding from their trial experience that the revision must be of higher quality than the original (despite being a product of their beliefs).

To summarize, our account holds that consumers will exhibit a preference for products that have been revised, even absent objective improvement. Specifically, we suggest that lay beliefs that revisers intend to improve their creations and that revisions represent the successful implementation of those intentions prompt consumers to place undue faith in revisions—to, in a sense, abandon critical assessment of revisions—leading to perceptions that revisions are superior to originals, even in cases where objective improvement is absent. This account also suggests that when prompted to consider the ulterior motives of revisers—thereby calling the lay belief into question—people’s uncritical eye may become more critical. More formally, we predict:

**H1a:** Relative to products that have not been revised, consumers exhibit a preference for products that have been revised.

**H1b:** This preference for products that have been revised can hold even when revised products are not objectively better.

**H2a:** This preference is driven by consumers’ inferences that revisers intended to improve the product.

**H2b:** In turn, these inferences of good intentions lead consumers to subject revised products to less critical assessment than their unrevised counterparts.

Accordingly,

**H3:** This preference is moderated by the reviser’s perceived trustworthiness: when consumers perceived the reviser to be untrustworthy, they will be less likely to prefer revised products relative to their unrevised counterparts.

Figure 1 summarizes our conceptual model.

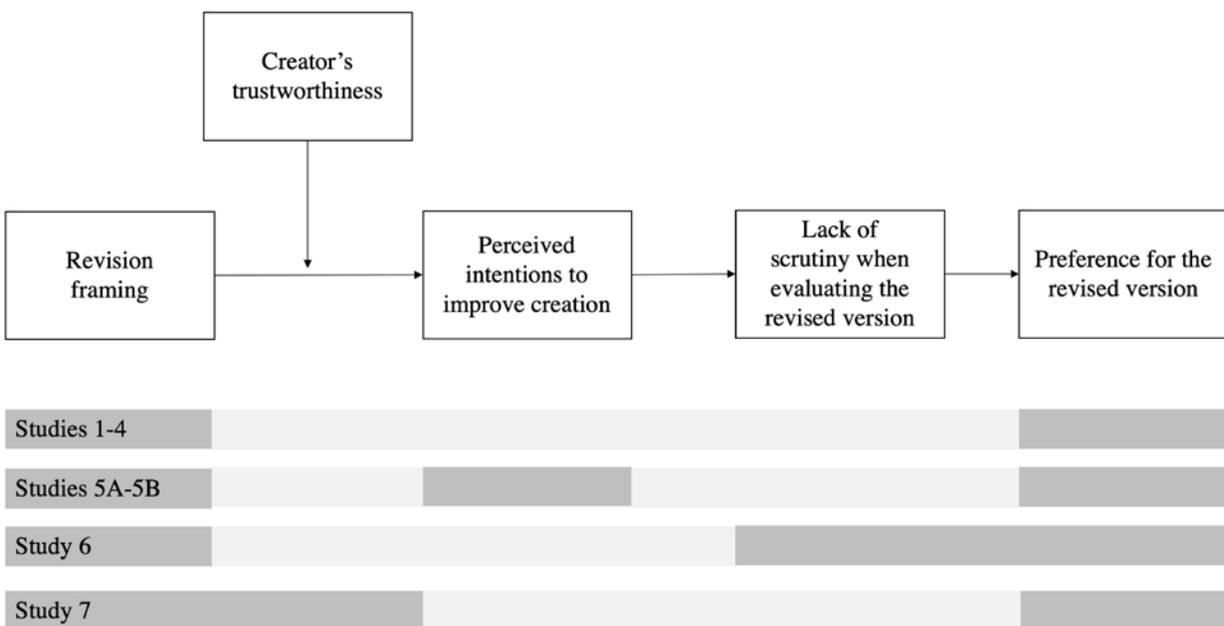


Figure 1. Conceptual Model.

## OVERVIEW OF STUDIES

Eleven studies conducted in the lab and field document a preference for revision, absent objective improvement, and highlight the role of overgeneralized inferences in this effect. Studies 1A-1C assess this preference among MBA students engaged in a resume revision process. Studies 2-3 then replicate the effect in the laboratory assessing experiential stimuli

(eating gummy candy) and naturalistic behavior (selecting a pen to sign a form). This preference for revised versions holds when the revision is trivial (i.e., does not represent improvement over the original, study 4A) and when there is no difference between the original and revised versions—when they are identical, save for the label (study 4B). Studies 5A and 5B show that the preference for revision is driven by inferences that revisers' intentions is to improve their creations; as a consequence, consumers are relatively uncritical of revisions' flaws (study 6). Finally, consistent with our proposition that consumers place undue faith in revisers' intentions, being informed that revisers are untrustworthy attenuates the preference for revision (study 7).

Across studies, our participants were diverse, drawn from student and community subject pools and national online panels. For online studies, we pre-specified the sample sizes to 75 to 250 per condition, depending on study design. For studies conducted in the laboratory, we pre-specified the sample sizes to at least 100 participants per condition. For study 1A, we collected data from all students who agreed to participate. For all studies, we report all manipulations and measures and did not analyze data until collection was complete. Unless otherwise indicated, no data were excluded and attrition rates were low (statistically equivalent across conditions and never exceeding 5%). Data and stimuli for all studies except study 1A are available via Open Science Framework: [https://osf.io/4x76c/?view\\_only=b3e050d5cb774780ac2a576b84660eca](https://osf.io/4x76c/?view_only=b3e050d5cb774780ac2a576b84660eca) (we cannot post data or resumes from study 1A because we do not have permission to do so from the university with which we collaborated).

### **STUDIES 1A-1C: MBA RESUME BLITZ**

Studies 1A-1C were three related studies using field and lab data. Study 1A took place during a process to help MBA students of a Northeastern US business school prepare their

resumes. MBA students submitted an initial version of their resume, were given opportunities to revise it based on feedback from resume coaches, and submitted a final version for inclusion in a “resume book” distributed to prospective employers. In study 1A, we focused on MBA students’ self-assessments of their resumes, predicting that the more extensive the revisions, the more positively they would rate their final resume.

In study 1B, we tested whether this perception is warranted: observers rated both versions of an MBA student’s resume; critically however, the two versions were simply presented as two “different” versions—i.e., observers were not told that one of the resumes was a revised version of the other, original, version. This design enabled us to measure whether the revisions were perceived as improved relative to the originals.

Then, in study 1C, we tested whether observers evaluated resumes more positively when they are labeled revised, irrespective of whether the given version had truly been revised. Half of observers simply rated the actual original, followed by its corresponding actual revision; these resumes were labeled correctly, as original and revised respectively. For the other half of observers, we swapped the labels: these participants first rated a version that was labeled “original” (that was actually the revision); they then rated a version that was labeled “revised” (that was actually the original). We predicted that participants would rate any versions labeled “revised” —regardless of whether they had truly been revised—to be better than those labeled “original,” demonstrating a preference for revision.

## **Methods**

**Study 1A. Creating and revising: do creators think their revisions are improvements?** We collaborated with the business school’s Career and Professional

Development (CPD) team, following their process to help students revise their resumes: students submitted an initial resume version to CPD. Over two months, students received feedback from a resume coach; students submitted a final resume for inclusion in a resume book distributed to prospective employers. This process granted us the opportunity to analyze a diverse set of genuinely revised stimuli.

In the fall of 2016, 302 students requested feedback. After students had submitted their final version, CPD asked whether they were willing to be contacted about the review process; we obtained the email addresses of the 77.1% of students who agreed (233 out of 302). We sent an email to these 233 students, hereafter referred to as “creators,” containing a link to a short survey (see web appendix for email text). A priori, we decided to send one follow-up email to non-responders and to close the survey once two full days had passed without any new survey responders, resulting in a response rate of 18.9% (44 out of 233).

In the survey, these 44 creators responded to four questions about the resume revision process. First, they rated, “What percentage of the final version of your resume is different from your original resume?” on a 0–100 scale with endpoints labeled 0 (*the final version is exactly the same as the original*) and 100 (*the final version is completely different from the original*). Second, they rated, “Relative to my original resume, my final resume is...” using the options: -3 (*dramatically worse*), -2 (*moderately worse*), -1 (*a bit worse*), 0 (*about the same*), 1 (*a bit better*), 2 (*moderately better*), and 3 (*dramatically better*). Third, they rated, “How satisfied are you with the final version of your resume?” on a scale from 1 (*not at all satisfied*) to 7 (*extremely satisfied*). Finally, they rated, “How many times did you obtain feedback from a resume coach?” via a numeric text entry box. This survey allowed us to assess whether creators indeed revised their resumes and viewed the final products as improved—which presumably is their goal in

having formally worked on their resumes over the course of the semester. At the end of the survey, we asked creators whether we could use their (anonymized) resumes for future research.

This process resulted in data from 33 creators (i.e., 33 pairs of resumes) to be used for Studies 1B and 1C. (In total, data from nine creators were excluded: five creators did not grant us permission to use their data and six creators did not answer all four survey questions).

**Study 1B. Are the revisions better than the originals?** Next, an independent sample of participants evaluated one of the resume pairs generated in study 1A. These observers ( $N = 204$ , 46.6% male;  $M_{\text{age}} = 35.58$  years,  $SD = 13.83$ ) were community members who came to the laboratory to participate in this and a series of unrelated studies. Observers rated both versions of one resume pair from study 1A (i.e., one MBA student's original resume and the same MBA student's revised resume; the name of each creator in all resumes was replaced with the generic name "Alex Newman" with all other author identifiers removed.). However, resumes did not have labels of original and revised. Instead, participants only read: "On the next pages, you will see two resumes and you will be asked to rate the appeal of each resume." Each resume was presented on a different page on a computer screen and participants were asked to rate the "overall appeal" on a scale from 1 (*very low*) to 7 (*very high*). Between-subjects, we manipulated the order in which the resumes were presented. Specifically, half of observers rated the original (though it was not labeled as original), followed by its corresponding revision (again, the revision was not labeled as such). The other half was presented with the versions in the opposite order (i.e., the revision first, followed by the original version, though they were not labeled as such). Thus, we ended up with a pool of 66 resume pairs (i.e., we had a pool of 33 pairs in study 1A where we counterbalanced which version was presented first).

**Study 1C. Are resumes evaluated more positively when labeled revised?** Finally,

another independent sample of participants evaluated the same resumes that observers in study 1B evaluated. These observers ( $N = 453$ , 50.1% male;  $M_{\text{age}} = 30.27$  years,  $SD = 12.27$ ), like those in study 1B, were community members who came to the laboratory to participate in this and a series of unrelated studies. As in study 1B, observers rated both versions of one resume pair (i.e., one MBA student's original resume and the same MBA student's revised resume); however, in this study, resumes had "original" or "revised" labels. First, observers were shown a version and asked to "please rate the ORIGINAL DRAFT of the resume (shown above) with respect to the following dimension: Overall appeal" on a 1 (*very low*) to 7 (*very high*) scale. On the next screen, observers were shown the other version in the given resume pair and asked to "please rate the REVISED DRAFT of the resume (shown above) with respect to the following dimension: Overall appeal" on a 1 (*very low*) to 7 (*very high*) scale.

Between-subjects, we manipulated whether the draft described as "original" was truly the original, and similarly, whether that described as "revised" was truly the revision. Specifically, half of observers simply rated the actual original, followed by its corresponding actual revision. The other half of observers first rated a version that was labeled "original" (but that was actually the revision); they then rated a version that was labeled "revised" (but that was actually the original). This set-up produced 66 resume pairs (i.e., a pool of 33 control pairs, in which actual originals were paired with their actual revisions; and a pool of 33 experimental pairs, in which actual revisions labeled as "original" were paired with their corresponding originals labeled as "revision"). Thus, each observer was randomized to rate both versions of a randomly selected resume pair from either the control pool or the experimental pool. Our results collapse across the 33 resumes within each pool. This and all subsequent studies concluded with basic demographic questions (e.g., gender, age, education level).

## Results

**Study 1A. Do creators think their revisions are improvements?** Creators deemed their revised resume to be significantly different from their original resume ( $M = 36.34\%$ ,  $SD = 26.33$ ; one sample  $t$ -test against 0%,  $t(37) = 8.51$ ,  $p < .001$ ), and deemed their revised resume to be significantly better than their original resume ( $M = 1.55$ ,  $SD = 0.90$ ; one sample  $t$ -test against the scale midpoint,  $t(43) = 11.38$ ,  $p < .001$ ). Moreover, the more dissimilar their two versions were, the higher quality the creators perceived their revised resume,  $r(38) = .612$ ,  $p < .001$ . This relationship held,  $b = 0.58$ ,  $SE = .004$ ,  $t(35) = 4.88$ ,  $p < .001$ , over and above the effect of the number of times the creator obtained feedback (which independently predicted perceptions of improvement,  $b = 0.37$ ,  $SE = .113$ ;  $t(35) = 3.07$ ,  $p = .004$ ). Creators were highly satisfied with their final products ( $M = 5.55$ ,  $SD = 1.00$ ). On average, they had obtained feedback 2.34 times ( $SD = 0.96$ ) over the course of the semester. Thus, these findings confirm that revised resumes differed from original resumes, and in the creators' eyes, should be noticed by others as significantly improved<sup>1</sup>.

**Study 1B. Are the revisions really any better than the originals?** When the resumes were simply presented as different versions—i.e., without the potentially value-laden labels of “original” and “revised”—observers perceived both resumes in a given pair as equally appealing ( $M_{\text{original}} = 5.08$ ,  $SD = 1.46$  vs.  $M_{\text{revised}} = 5.00$ ,  $SD = 1.43$ ). A  $2 \times 2$  mixed ANOVA revealed that revised resumes were seen as no better than original resumes,  $F(1, 202) = 0.99$ ,  $p = .321$ ,  $\eta_p^2 = .005$ ; similarly, resumes presented on the first screen were considered equally appealing as

---

<sup>1</sup> Six creators in study 1A did not answer the following question: “What percentage of the final version of your resume is different from your original resume?” Thus, for analyses involving this measure, the valid sample is 38 participants.

resumes presented on the second screen (i.e., no effect of presentation order:  $F(1, 202) = 0.32, p = .572, \eta_p^2 = .002$ ; no significant interaction:  $F(1, 202) = 1.72, p = .191, \eta_p^2 = .008$ ).

**Study 1C. Are resumes evaluated more positively when labeled revised?** Observers exhibited a preference for resumes labeled as “revisions,” independent from the veracity of this label. Specifically, a  $2 \times 2$  mixed ANOVA revealed that while resumes accurately labeled as revised were seen as no better than resumes accurately labeled as originals,  $F(1, 451) = 0.10, p = .754, \eta_p^2 = .000$ , observers perceived resumes labeled as revisions more positively than those labeled as originals,  $F(1, 451) = 51.61, p < .001, \eta_p^2 = .103$ —an effect that held even when originals were *merely labeled* as revisions,  $t(226) = 4.20, p < .001, d = 0.29$  (see Figure 2).

In sum, studies 1A-1C suggest that creators confound change with improvement: the more they changed their resumes, the more they thought their resumes had improved (study 1A). However, these assessments are inaccurate: observers do not find the revisions to be any better than the originals (study 1B). In fact, the only thing that made observers perceive improvement was the mere revision label—irrespective of whether the given version had truly been revised (study 1C).

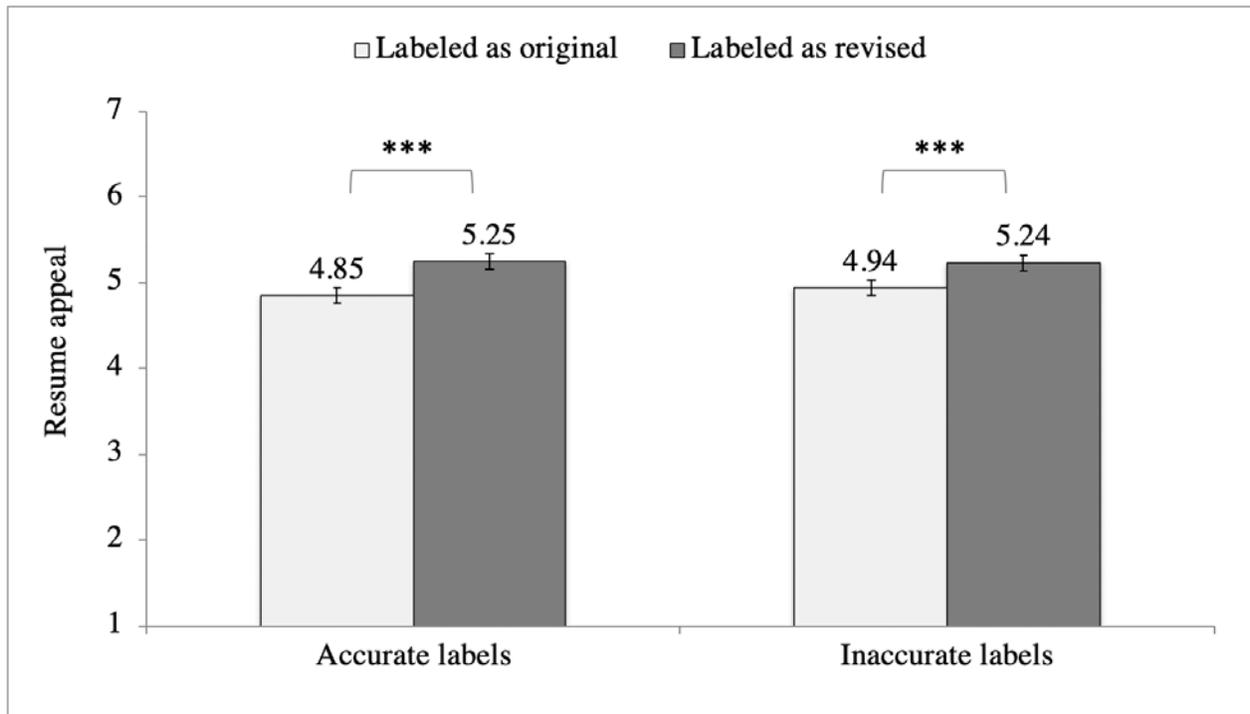


Figure 2. Study 1C: mean ratings of resumes across conditions. Error bars indicate  $\pm 1$  SEM.  $*p < .05$ ,  $**p < .01$ ,  $***p < .001$ .

## STUDY 2: EATING GUMMIES

Study 2 has two primary goals. First, it tests whether this preference for revision generalizes to an experiential stimulus (eating gummies). Second, given past research that highlights a tendency to reward effort (Buell and Norton 2011; Chinander and Schweitzer 2003; Diehl, Van Herpen, and Lambertson 2015; Klein and O'Brien 2017; Kruger, Wirtz, Van Boven, and Altermatt 2004; Morales 2005), study 2 also assesses whether this preference for revision emerges over and above a preference for effort.

### Methods

Participants ( $N = 239$ ; 49.8% male;  $M_{\text{age}} = 23.19$  years,  $SD = 4.50$ ) were community

members who came to a behavioral lab to participate in this and a series of unrelated studies. Each participant was seated at her own private computer cubicle. Two cups were placed on each desk, each containing one of two similar but distinct gummy candies (see photographs of study stimulus in web appendix).

Participants followed instructions that appeared on successive computer screens. First, they were told that they would sample two gummies made by different companies. Participants were told that the gummies cost and took about the same time to make. Next, participants ate the gummy on the left, which served as the control gummy, and indicated their agreement with three statements: this candy is high quality, this candy is tasty, and the company put a lot of effort into making this candy, each on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Unlike study 1C, in this study the control stimulus was not given a special label—i.e., it was not labeled as “original.”

Next, participants were told that the other gummy—the gummy on the right, which served as the experimental gummy—had been revised. They read: “...the recipe for this candy is the final, market-ready version of the recipe. Before the recipe was finalized, the company had tried several different recipes before finalizing the current formulation.” Participants then ate this gummy and completed the same three items as the control gummy. At the end of the study, participants indicated whether they had sampled both gummies (i.e., whether they had complied with the instructions). Importantly, the revision label was arbitrary: between-subjects, we counterbalanced which of the two gummies was placed on the left versus right and whether participants ate and rated the control or the revised gummy first.

## **Results**

Again, participants exhibited a preference for the revised option. A composite measure of the two primary outcome measures (quality and taste,  $r_{\text{control}} = .71$ ,  $r_{\text{experimental}} = .78$ ) revealed that the revised gummy was deemed superior to the control gummy ( $M_{\text{experimental}} = 3.77$ ,  $SD = 1.77$ ;  $M_{\text{control}} = 3.47$ ,  $SD = 1.57$ ),  $t(238) = 2.41$ ,  $p = .017$ ,  $d = 0.16$ —despite the fact that the gummies were *merely labeled* this way (effects held when the measures are analyzed separately, see web appendix). This effect of revision label held when controlling for the difference in perceived effort in creating the revised versus control gummy ( $F(1, 237) = 4.10$ ,  $p = .044$ ,  $\eta_p^2 = .017$ ). This is important because the company was perceived to have exerted greater effort in creating the revised gummy ( $M_{\text{experimental}} = 3.99$ ,  $SD = 1.80$ ;  $M_{\text{control}} = 3.41$ ,  $SD = 1.63$ ),  $t(238) = 4.94$ ,  $p < .001$ ,  $d = 0.32$ ). Finally, the results are also substantively equivalent when excluding the 22 participants who reported that they did not sample both gummies (see web appendix).

In sum, study 2 replicates this preference for revision using a stimulus that gave participants the entire bottom-up experience (i.e., actually eating the gummies) before making their evaluations. A gummy framed as revised tasted better than a control gummy—an effect that emerged over and above perceptions of increased effort.

### **STUDY 3: PEN CHOICE**

In Studies 1C and 2, observers perceived resumes and candies arbitrarily labeled as revised to be of relatively high quality. These findings are consistent with our proposition that consumers exhibit a preference for revision, absent objective improvement. However, there is a possibility that these effects are partly a product of demand, whereby participants gave favorable ratings to stimuli labeled “revised” not because they were truly impressed with the revisions, but because they thought the experimenters wanted them to do so. Therefore, in study 3, we address

this possibility by making our test of a preference for revision peripheral to the ostensible task of the study; the study was conducted surreptitiously, during the introductory paperwork that participants completed at the start of a battery of lab studies. Specifically, participants were asked to sign an “administrative form,” which detailed the lab’s procedures. We measured which of two pens they used to sign the form, one of which bore a revision label. This study was pre-registered: AsPredicted.org (#23900).

## **Methods**

Participants ( $N = 204$ ; 46.6% male;  $M_{\text{age}} = 35.58$  years,  $SD = 13.83$ ) were community members who came to a behavioral lab of a Northeastern US University to participate in a series of unrelated studies. Upon arrival, each participant was seated at their own private computer cubicle. Immediately after participants provided consent on the computerized form, they were informed that before starting with the studies, they had to review and sign a paper administrative form describing the lab policies. Two pens were lying on each desk; participants used whichever one they wanted to sign the form (see photographs of the setup in web appendix).

We manipulated both the color and revision label of the pens: one pen was black, the other was blue; and, critically, one pen was labeled as having been revised, whereas the other bore no such label. Between-subjects, we counterbalanced the ascription of color to revision status (i.e., half of the participants chose between a blue pen and a revised black pen; the other half chose between a black pen and revised blue pen). Color and revision status was conveyed to participants via the label that was printed directly on the pen (see photographs of these pens in web appendix, which we had made specifically for this study). One of the pens had a standard label—i.e., it simply was labelled with its color; the other had a label that in addition to denoting

its color, also conveyed that it had recently been revised. Specifically, participants faced a choice between either a blue pen labeled “BLUE” and a black pen labeled “BLACK (revised 2019)” or a black pen labeled “BLACK” and a blue pen labeled “BLUE (revised 2019). Between-subjects, we also counterbalanced the position of the pens on the desk (for half of participants, the revised pen was on the left; for the others, the revised pen was on the right).

After signing the form, participants completed the surveys on their computer. The first question was an initial exploratory test of the idea that revision labels may cause consumers to not put too much thought into their choices, we asked participants “How did you make this choice?,” offering the response options: “I didn't think about the choice and just picked up one of the two pens available” and “I thought about the choice before selecting one of the two pens available” (we return to this idea in studies 6-7, where we more directly and systematically test it).

We also assessed their pen color choice by asking participants to answer this second question: “What color pen did you choose?” and offering the response options: “blue” or “black” (participants also entered a unique ID number at the start of the battery of surveys, which enabled us to match the color choice with whether it had a revised label for the given participant). At the end of the session, we debriefed participants, informing them that the administrative form was part of this experiment, in which we tested for a preference for revision.

## **Results**

As predicted, participants showed a preference for the pen labeled “revised,” with 56.9% of participants choosing it over the pen with the standard label; this percentage was significantly different from 50%,  $\chi^2(1) = 3.84, p = .050$ . This study also provides suggestive evidence that is

consistent with our process account. Specifically, we posit that perceivers assume that revisions are the fruition of an intention to improve the creation, which, if true, might lead consumers to be less likely to carefully scrutinize something that has been labeled “revised.” Consistent with this account, 65.7% of participants said that they “didn’t think about the choice” compared to 34.3% who said they thought about it ( $\chi^2(1) = 20.08, p < .001$ ).

## **STUDIES 4A-4B: RESUMES AND LOGOS**

Studies 1-3 support our proposition that consumers prefer products that have been revised, absent objective improvement. Studies 4A and 4B offer even more conservative tests of this proposition. Study 4A tests whether the preference emerges when the difference between the original and the revised version is trivial, and this triviality is transparent. Study 4B goes further, testing whether the preference for revision arises when in fact there is *no* difference between the versions—i.e., when the original and the revised versions are identical, save for the label.

### **Study 4A**

In study 4A, participants rated two resumes differing only in font: an original and a revised version. We assessed this preference for revision at a perceptual level, examining whether the attributes of a font (e.g., readability, clarity) are perceived differently as a function of revision label.

### **Methods**

Participants ( $N = 401$  MTurk workers; 50.6% male;  $M_{\text{age}} = 37.40$  years,  $SD = 12.10$ ) completed this study, which was the second of two unrelated studies administered together for a

fixed payment. Participants were shown two versions of a person's resume differing only in font (Athelas vs. Lao); resumes were presented on the same page so participants could clearly compare both versions and verify that the only change between them was font (see resumes used in web appendix). One version was labeled "original;" the next, "revised." To ensure the arbitrariness of the revision, between-subjects, we manipulated the ascription of revision label to resume: for half of participants, the Athelas resume was labeled "original" and the Lao resume was labeled "revised"; these ascriptions were reversed for other participants.

Immediately below the resumes, participants rated the appeal of each version: as in study 1C, participants were asked to "please rate the ORIGINAL DRAFT of the resume with respect to the following dimension: Overall appeal" and to "please rate the REVISED DRAFT of the resume with respect to the following dimension: Overall appeal," on a 1 (*very low*) to 7 (*very high*) scale. Below this primary outcome measure, participants indicated which font was: clearer, more professional, more readable, nicer, and more appealing by selecting one of three response options: the font of the original version, the font of the revised version, or the two fonts are about the same. Thus, this study did not require a forced choice between the original and revised version. Instead, participants were given a third alternative: they could indicate that both the original and the revised options were the same.

## **Results**

A  $2 \times 2$  mixed ANOVA on overall appeal revealed the critical main effect of revision label,  $F(1, 399) = 29.65, p < .001, \eta_p^2 = .069$ : whatever resume was merely labeled "revised" was perceived as more appealing than the one labeled "original" (see Figure 3). We again observed that participants preferred the revised option, here in a case when the "revision" was

obviously trivial. There was also a main effect of the ascription of label to resume,  $F(1, 399) = 13.29, p < .001, \eta_p^2 = .032$ , indicating that resume appeal was higher for one ascription (i.e., Original: Athelas; Revised: Lao, two left-most bars of Figure 3) than the other, but this main effect was incidental as it did not interact with revision label (i.e., the revision boost was equally strong across ascriptions),  $F(1, 399) = 1.72, p = .190, \eta_p^2 = .004$ .

Ratings of font attributes provide further evidence for this preference for revision. Holding font constant, a given font was perceived as better on all measured attributes—clarity, professionalism, readability, niceness, and appeal—when it was used in the version labeled “revised” relative to the version labeled “original” (see Figure 4); a two-sample z-test assessed whether the percentage of participants who preferred the font used on the original version was significantly different from the percentage of participants who preferred the font used on the revised version for each attribute (all  $ps < .001$ ).

In sum, study 4A shows that the preference for revision holds when the revision is obviously trivial, and this triviality was made salient by presenting both resumes on the same page. Study 4A also suggests that this preference operates at a perceptual level: the fonts were perceived to have different attributes as a function of the (arbitrary) revision label.

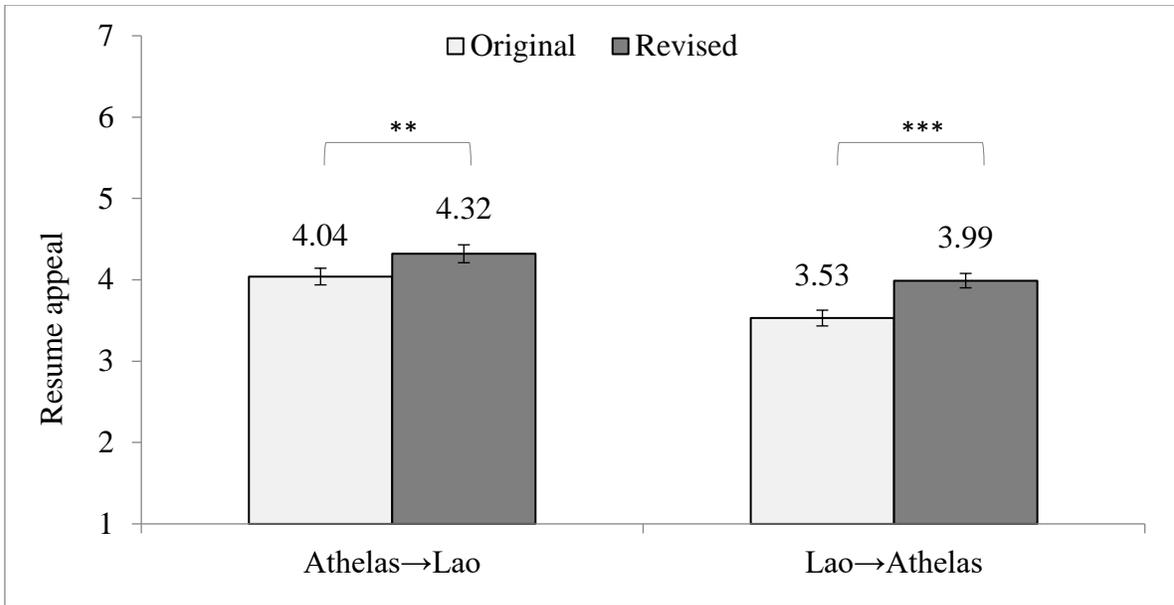


Figure 3. Study 4A: mean ratings for original and revised versions across conditions. Error bars indicate  $\pm 1$  SEM. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

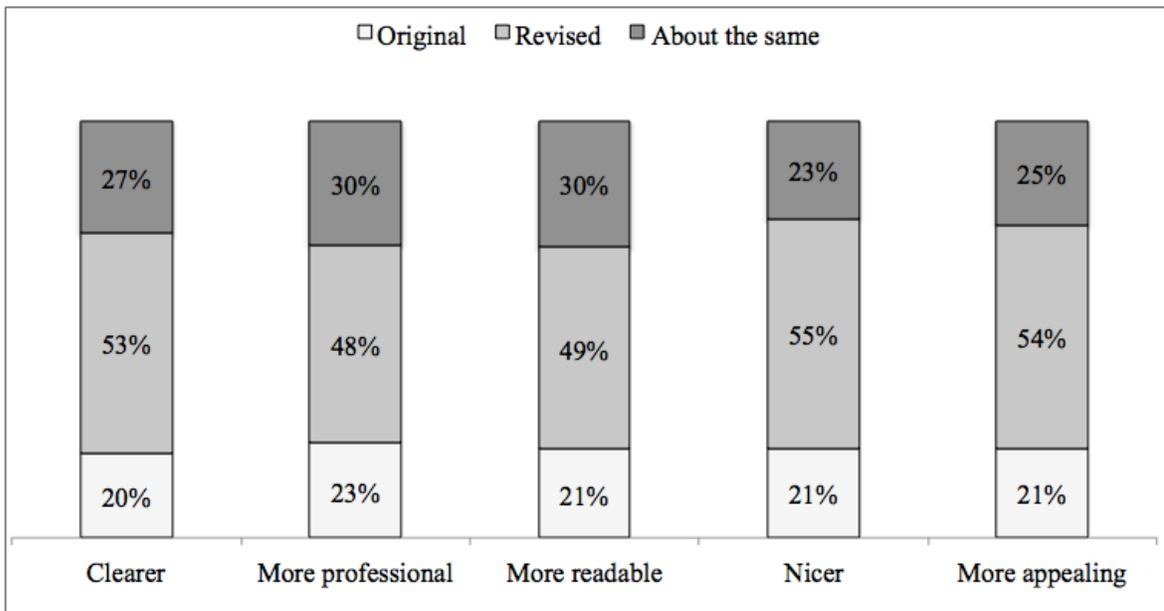


Figure 4. Study 4A: participants' assessments of which font was superior for each of five attributes. The figure plots, for each attribute, the proportion of participants indicating that the

font on the original version was superior (white above); the font on the revised version was superior (light grey above); or that the neither font was better than the other (dark grey above).

## **Study 4B**

Study 4B tests whether the preference for revision arises when in fact there is *no* difference between the versions—i.e., when the original and the revised versions are identical, save for the label. In study 4B, participants rated three versions of a logo: for everyone, the first two logos were labeled as original and revised versions. In the control condition, the third logo was labeled as “back to original” while in the experimental condition the third logo was labeled as “2<sup>nd</sup> revised version”. Our primary prediction was that the 3<sup>rd</sup>-presented logo would be more appealing when labeled “2<sup>nd</sup> revised version” relative to when labeled “back to the original version.” Because this contrast is restricted to the 3<sup>rd</sup>-presented logo, it controls for mere exposure (Bornstein 1989; Zajonc 1968): in both conditions, participants rated a logo that they had seen before (i.e., the third logo had already been presented as the 1<sup>st</sup> logo).

## **Methods**

Participants ( $N = 409$  Mturk workers; 59.4% male;  $M_{\text{age}} = 34.48$  years,  $SD = 10.47$ ) completed this study, which was the first of three unrelated studies administered together for a fixed payment. Participants rated the appeal of different versions of a logo for “Cleansy” (a fictitious soap) across three presentations, using the same scale as in study 4A (overall appeal). Participants first saw and rated the appeal of one of the logos, labeled “original version.” This logo served as the focal logo; it would be later presented a second time. On the next screen, participants were shown and rated a different logo, labeled “revised version.” On the third

screen, participants were shown and rated the focal (i.e., first-presented) logo again. Between-subjects, we manipulated the revision label on this third screen: in the control condition, it was labeled “back to the original version;” in the experimental condition, it was labeled “2<sup>nd</sup> revised version” (see Table 1 for the logos and overview of the design). To ensure the arbitrariness of revision status, we manipulated the ascription of logo to focal logo status, between-subjects. At the end of the study, participants answered a reading check question assessing their knowledge of the number of logos presented.

*Table 1. Overview of design*

	1 <sup>st</sup> -presented logo (focal logo)	2 <sup>nd</sup> -presented logo (a different logo)	3 <sup>rd</sup> -presented logo (focal logo again)
			
Control	“original version”	“revised version”	“back to the original version”
Experimental	“original version”	“revised version”	“2 <sup>nd</sup> revised version”

*Note.* This table collapses across the ascription of logo to focal logo status.

## Results

A comparison of the appeal of the 3<sup>rd</sup>-presented logo between conditions tests our primary prediction. A  $2 \times 2$  between-subjects ANOVA controlling for logo order revealed a main effect of revision label: the third logo was more appealing when it was labeled “2<sup>nd</sup> revised version” ( $M_{\text{experimental}} = 3.54$ ,  $SD = 1.46$ ) relative to when labeled “back to the original version” ( $M_{\text{control}} = 3.26$ ,  $SD = 1.43$ ,  $F(1, 405) = 3.91$ ,  $p = .049$ ,  $\eta_p^2 = .010$ ). Thus, this preference for revision held even when the revision was literally identical to an original version, from an

identical point in time, but merely labeled “revision.” Also, the effect of revision happened to be stronger when one of the labels was assigned to focal label status over the other (i.e., interaction:  $F(1, 405) = 3.88, p = .050, \eta_p^2 = .009$ ), but there was no main effect of logo order,  $F(1, 405) = 0.46, p > .250, \eta_p^2 = .001$ . Results hold when excluding participants who did not pass the reading check question (7.6%; see analysis in web appendix).

As a conceptual replication of the basic effect, we compared the appeal of the 1<sup>st</sup> versus 2<sup>nd</sup>-presented logo. A  $2 \times 2$  mixed ANOVA revealed a main effect of revision label,  $F(1, 407) = 36.08, p < .001, \eta_p^2 = .081$ : the 2<sup>nd</sup>-presented logo, which was labeled “revised version” ( $M_{\text{revised}} = 3.42, SD = 1.36$ ) was more appealing relative to the 1<sup>st</sup>-presented logo, which was labeled “original version” ( $M_{\text{control}} = 3.10, SD = 1.41$ ). There was a marginally significant main effect of logo,  $F(1, 407) = 2.93, p = .088, \eta_p^2 = .007$ , but this factor did not interact with revision label,  $F(1, 407) = 0.32, p > .250, \eta_p^2 = .001$ .

## **STUDIES 5A-5B: POETRY**

Studies 1-4 provide converging evidence of a preference for revision, absent objective improvement. Studies 5A and 5B provide evidence of why this effect occurs. Specifically, we test whether it is driven by inferences about the reviser’s intentions: consumers infer that the reviser sincerely wanted to improve the product, and that the revision represents the successful fruition of those intentions. Study 5A provides preliminary evidence of mechanism by examining whether a belief that the reviser was genuinely trying to improve their creation is positively correlated with the propensity to see the revision as an improvement relative to the original. Next, study 5B tests this mechanism via moderation, experimentally manipulating the reviser’s intentions, and assessing whether this factor interacts with revision labelling.

## Study 5A

### Methods

Participants ( $N = 300$  Mturk workers; 57.7% male;  $M_{\text{age}} = 36.09$  years,  $SD = 11.27$ ) completed this study for a fixed payment. Participants were instructed to evaluate the quality of the original and the revised version of a poem written by “Riley, a freelance writer hired by a company that creates and sells books of short poems and stories.” We had two versions of a poem with three stanzas each (see poems used in web appendix). As in studies 4A and 4B, the revision label was arbitrary: between-subjects, we manipulated the ascription of revision label to poem: for half of participants, poem A was labeled “original” and poem B was labeled “revised”; these ascriptions were reversed for other participants. First, participants saw the original poem and rated its overall quality (1 = very low to 7 = very high). On the next page, participants saw and rated the overall quality of the revised poem. On the third page, participants answered one question to assess the creator’s intentions: “To what extent do you think that Riley, the freelance writer, genuinely tried to make the poem better during the revision process?” (1 = not at all to 7 = very much).

### Results

A  $2 \times 2$  mixed ANOVA on overall quality of the poem revealed the main effect of revision label,  $F(1, 298) = 18.82, p < .001, \eta_p^2 = .059$ : poems labeled as revised were rated higher ( $M_{\text{revised}} = 4.90, SD = 1.41$ ) than poems labeled as original ( $M_{\text{original}} = 4.50, SD = 1.30$ ). We did not observe a significant effect of poem order ( $F(1, 298) = 0.03, p = .858, \eta_p^2 = .000$ ) or a significant interaction between label and poem order ( $F(1, 298) = 0.71, p = .402, \eta_p^2 = .002$ ).

Importantly, we found that to the extent that participants believed that the creator had genuine intentions during the revision process, they thought the revised poem was of higher quality relative to the original poem. The SPSS MACRO MEMORE (Montoya 2019) indicated that perceptions of the creator's intentions were positively associated with higher ratings of the revised version relative to the original version ( $b = 0.60$ ,  $SE = 0.06$ ,  $p < .001$ ). Specifically, participants preferred the revised poem to the original poem to the extent that their motive ratings were roughly at or above the midpoint of the scale (Johnson-Neyman point: 3.96).

### **Study 5B**

Study 5B was a  $2 \times 3$  mixed design in which we manipulated revision label (original vs. revised, within-subjects) and the reviser's motive (control vs. explicit improvement motive vs. explicit money motive, between-subjects). We predicted an interaction such that when told that the reviser's motive was to make money, the preference for revision would be dampened. And, if participants naturally infer revisers to have an improvement motive, then participants in the control condition—who were not given any information about the creator's motive to revise—should, like those in the “improvement motive” condition, be impressed with the revision.

### **Methods**

Participants ( $N = 600$  Mturk workers; 45.8% male;  $M_{\text{age}} = 36.80$  years,  $SD = 11.09$ ) completed this study for a fixed payment. Like in study 5A, participants were asked to evaluate the quality of the original and revised versions of a poem written by Riley, a freelance writer hired by a company that creates and sells books of short poems and stories.

Participants were first presented with a poem labeled “original version,” and rated its overall quality on a scale from 1 (very low) to 7 (very high). Then, they were presented with a

poem labeled “revised version,” and rated its overall quality on the same 7-point scale. As in study 5A, the revision label was arbitrary: between-subjects, we manipulated the ascription of revision label to poem (this factor interacted with label; however, results held regardless of which poem had the revision label and therefore in the results section we collapse across this third factor).

Prior to viewing the revised poem, all participants were told: “Riley revised the poem.” Next, we gave some participants information about why Riley revised the poem. Specifically, in the *improvement motive* condition, participants were told: “Riley revised the poem because she sincerely wants to deliver a poem of high quality.” In the *money motive* condition, participants were told: “Riley revised the poem because she wants to make money. She is paid an hourly rate, so the more time she spends working on the poem, the more money she makes. She doesn’t particularly care about whether the poem is high quality.” And in the control condition, participants were not given any information about Riley’s motives. After rating both poems, participants answered a reading check question (“Based on the information you were given in this study, why did Riley revise the poem?,” given forced-choice options of each condition).

## Results

There was a main effect of revision label: collapsing across the motives manipulation, the poem label as revised was deemed to be of higher quality than the original poem ( $M_{\text{original}} = 4.46$ ,  $SD = 1.17$  vs.  $M_{\text{revised}} = 4.85$ ,  $SD = 1.38$ ;  $F(1, 597) = 44.03$ ,  $p < .001$ ,  $\eta_p^2 = .069$ ). There was also a main effect of motives ( $F(2, 597) = 6.87$ ,  $p = .001$ ,  $\eta_p^2 = .023$ ). However, these main effects were qualified by an interaction,  $F(2, 597) = 4.07$ ,  $p = .018$ ,  $\eta_p^2 = .013$  (see Figure 5). Specifically, participants in the control and improvement motive conditions exhibited a

preference for revision (control:  $M_{\text{original}} = 4.63$ ,  $SD = 1.09$  vs.  $M_{\text{revised}} = 5.06$ ,  $SD = 1.31$ ;  $F(1, 597) = 16.31$ ,  $p < .001$ ,  $\eta_p^2 = .027$ ; improvement motive:  $M_{\text{original}} = 4.39$ ,  $SD = 1.19$  vs.  $M_{\text{revised}} = 4.96$ ,  $SD = 1.40$ ;  $F(1, 597) = 32.55$ ,  $p < .001$ ,  $\eta_p^2 = .052$ ). However, this preference was smaller and only marginally significant in the money motive condition ( $M_{\text{original}} = 4.37$ ,  $SD = 1.20$  vs.  $M_{\text{revised}} = 4.55$ ,  $SD = 1.37$ ;  $F(1, 597) = 3.00$ ,  $p = .084$ ,  $\eta_p^2 = .005$ ). These results hold when excluding participants who did not pass the reading check question (13.3% - see analysis in web appendix).

Taken together, Studies 5A and 5B provide converging evidence that the preference for revision is driven by inferences about the reviser's intentions. Study 5A shows that the tendency to view revisions as appealing is positively associated with the belief that the reviser was sincerely trying to improve their creation. Study 5B goes further, showing that the preference is moderated by the reviser's motives. Specifically, when participants were told that the reviser's motive was not to improve their creation per se, but rather, to make money, the preference for the revision is dampened—despite the fact the poems themselves were identical across the motives manipulation. And, attesting to our proposition that consumers tend to assume that revisers want to improve the creation, the preference for revision was as strong in the control condition as it was in the improvement motives condition.

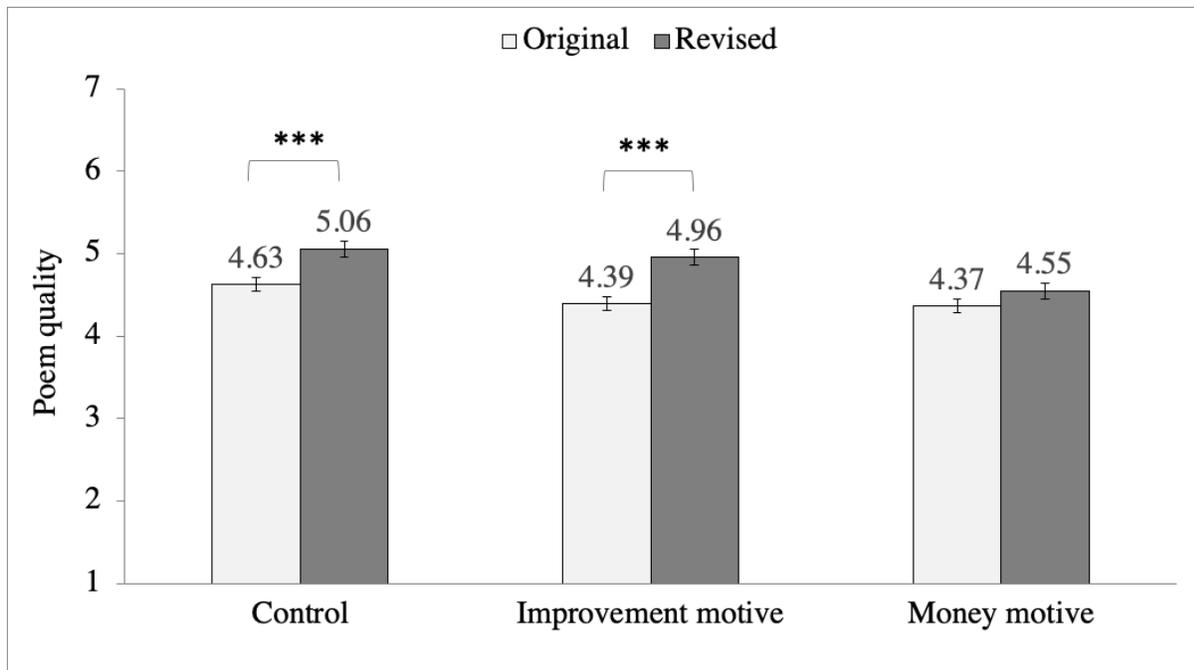


Figure 5. Study 5B: mean ratings for original and revised poems across conditions. Error bars indicate  $\pm 1$  SEM. Significance levels compare the two versions within each condition ( $*p < .05$ ,  $**p < .01$ ,  $***p < .001$ ).

## STUDY 6: BUGGY VIDEO GAME

So far, we have shown that consumers have a preference for revision, absent objective improvement, and that this effect is driven by inferences about the reviser's intentions—consumers infer that the reviser sincerely wanted to improve the creation, and that the revision realizes those intentions. In study 6, we test our fourth proposition: as a consequence of these inferences, consumers may place undue faith in the revision, causing them to evaluate the revision with reduced scrutiny. Specifically, study 6 tests whether consumers are less likely to detect flaws in a product when told that the product has been revised.

Study 6 also tests for the robustness of the basic effect. First, it tests whether the preference for revision manifests in a between-subjects design using the same single experiential

stimulus—playing a video game—while merely framing the stimulus as more, versus less, revised. Second, it tests whether the effect holds when we invoke revision without using the word “revision” —here, we denote revision by version number (e.g. “Version 5.0”). Finally, it also tests whether the effect manifests with a conservative manipulation of revision status: here, we manipulate the degree of revision (as opposed to the presence versus absence of revisions).

Study 6 was a two condition between-subjects design: in the *less-revised* condition, the video game was referred to as version 2.0; in the *more-revised* condition, it was referred to as version 5.0. In fact, the video game was identical across conditions: all participants played the same video game, and so all were endowed with the same bottom-up information about the game, experienced in full first-hand. Importantly, we specifically instructed the programmer to build subtle, ambiguously “buggy” features into the game—nothing that obviously disrupted the game, but that could be interpreted as an unintended bug (e.g., the cursor would sometimes lag a split-second behind). Although revision status should not bear on evaluating the game’s judged quality (because participants can simply rely on their actual experience playing it), we predicted that participants who think they are playing the more-revised version may be less likely to encode the ambiguous bugs as bugs—thereby creating a more enjoyable experience. This study was pre-registered: AsPredicted.org (#27218).

## **Methods**

Participants ( $N = 500$  Mturk workers; 53.8% male;  $M_{\text{age}} = 36.02$  years,  $SD = 11.14$ ) completed this study for a fixed payment. To begin, all participants read that we—the requesters of the Mturk HIT—have been working on developing a game called *ART Time*, and were shown a brief description of the game. The game allows users to freely “paint” a blank canvas, with a

selection of different tools that create different shapes and colors. We hired a developer to create the game specifically for this study, ensuring that all participants would objectively have the same novel experience. Moreover, we instructed the developer to design the game with subtly, ambiguously “buggy” features (e.g., the cursor would sometimes lag a split-second behind).

Between-subjects, all participants read that, to date, we had released five updates to the game. Participants in the *more-revised* condition were informed that they had been randomized to play Version 5.0, whereas those in the *less-revised* condition were informed that they had been randomized to play Version 2.0. In both conditions, we described the given version as having been developed in “early 2019” to hold constant participants’ knowledge of chronological recency. We did not show any cover art or other previews of the game so as to hold constant objective expectations.

Next, all participants played the game for two minutes, during which all other keyboard controls were disabled. When time expired, the page automatically continued to a survey screen where participants rated a block of enjoyment questions and a bugginess question (the order of the enjoyment block and the bugginess question was randomized, each presented on individual pages). In the enjoyment block (dependent variable), participants rated five items each on a 1 (*not at all*) to 7 (*extremely*) scale: how much they liked the game, how fun, enjoyable, and cool it was, and how happy they were playing it (hereafter referred to as the game enjoyment scale;  $\alpha = .97$ ). In the bugginess question, participants indicated the number of “specific individual bugs” they felt they had experienced in the game, from 0 to 20 (with a 21<sup>st</sup> option, “More than 20 [please type your number]”). Afterwards, participants who reported noticing any number of bugs greater than 0 were asked to provide some examples, via an open-ended essay box. Finally, all participants reported any general confusion with the task (99.6% of participants reported no

confusion) and answered a reading check in which they identified which version (of five) they had played.

## Results

Three critical findings emerged. First, participants again exhibited a preference for revision: an independent samples *t*-test on enjoyment found that participants enjoyed their experience more when told they were playing Version 5.0 ( $M_{\text{more-revised}} = 5.85$ ,  $SD = 1.29$ )—relative to Version 2.0 ( $M_{\text{less-revised}} = 5.26$ ,  $SD = 1.50$ ,  $t(498) = 4.75$ ,  $p < .001$ ,  $d = 0.42$ —even though all participants played the same game.

Second, consistent with this preference, participants noticed fewer bugs when told they were playing Version 5.0 ( $M_{\text{more-revised}} = 1.75$ ,  $SD = 4.07$ ) compared to Version 2.0 ( $M_{\text{less-revised}} = 2.85$ ,  $SD = 4.13$ ),  $t(498) = -2.98$ ,  $p = .003$ ,  $d = -0.27$ . Common descriptions of these alleged bugs included perceived speed issues (e.g., “It lagged a bit overlaying other stroke patterns”; “It did not move where I wanted it to”) and a perceived lack of features (e.g., “unable to erase”; “no choice of color, limited control”). Interestingly, significantly more participants noticed zero bugs in the Version 5.0 condition (65.1%) compared to those in the Version 2.0 condition (40.7%),  $\chi^2(1, N = 500) = 29.76$ ,  $p < .001$ ,  $\phi = 0.24$ . Results hold when excluding participants who did not pass the reading check question at the end (5.2% - see analysis in web appendix).

Third, the effect of revision framing on game enjoyment was mediated by number of bugs noticed while playing the game (PROCESS Model 4 with 5,000 bootstrapped samples). When entering number of bugs noticed to the model, the effect of revision framing on game enjoyment was significantly reduced (from  $b = 0.59$ ,  $SE = 0.6$ ,  $p < .001$  to  $b = 0.49$ ,  $SE = 0.12$ ,  $p$

< .001). More importantly, we found a significant indirect effect via number of bugs noticed ( $ab = 0.10$ , Boot  $SE = 0.05$ , 95%  $CI_{boot} [0.03, 0.22]$ ).

In sum, in study 6, a mere revision label changed how consumers experienced an otherwise identical game: participants who thought they were playing the revised version were less critical of the experience and therefore ended up liking the game more. Moreover, study 6 speaks to the robustness of the effect: it manifests in a between-subjects design in which we manipulated the degree of revision, denoted by version number. Finally, this study further addresses the possibility that this preference for revision is an artifact of mere exposure because all participants played the same game for the same amount of time (i.e., we held constant the amount of exposure).

## **STUDY 7: ARCHITECTURAL DRAWINGS**

Consistent with the idea that consumers place undue faith in the purity of reviser's intentions, in study 7, we assess whether the preference for revision is lessened when observers receive information about the reviser's (un)trustworthiness. Study 7 was a  $2 \times 2$  mixed design in which we manipulated revision label (original vs. revised, within-subjects) and the reviser's trustworthiness (low vs. high, between-subjects). We predicted an interaction such that when the reviser is untrustworthy, participants would be less likely to perceive the revision to be of relatively high quality.

### **Methods**

Participants ( $N = 400$  Mturk workers; 44.3% male;  $M_{age} = 37.95$  years,  $SD = 11.25$ ) completed this study for a fixed payment. Participants were asked to imagine that they were renovating their house and that their general contractor hired an architect to redesign the exterior.

First, all participants rated the overall appeal of the original draft of the architect's drawing, on a 1 (*very low*) to 7 (*very high*) scale (see drawings used in web appendix). On the next page, participants read that they had looked up the architect's trustworthiness on a local ratings website and discovered that this particular architect had a trustworthiness rating of 4.5/5.0 (*high trust* condition) or of 1.5/5.0 (*low trust* condition). Next, participants read that the architect revised the drawing, were shown the revised version, and rated its overall appeal on a 1 (*very low*) to 7 (*very high*) scale. As in previous studies, the ascription of revision status to drawing was counterbalanced between-subjects (this factor had no effect; therefore in the results we collapse across it). Finally, as a manipulation check, participants rated the architect's trustworthiness.

## Results

Participants in the high trust condition found the architect more trustworthy ( $M_{\text{high}} = 4.29$ ,  $SD = 0.63$ ) than participants in the low trust condition ( $M_{\text{low}} = 2.13$ ,  $SD = 0.92$ ,  $t(398) = 27.35$ ,  $p < .001$ ,  $d = 2.74$ ), suggesting that the trustworthiness manipulation was successful.

A  $2 \times 2$  mixed ANOVA revealed a main effect of trust ( $F(1, 398) = 14.96$ ,  $p < .001$ ,  $\eta_p^2 = .036$ ) and no main effect of label ( $F(1, 398) = .03$ ,  $p = .865$ ,  $\eta_p^2 = .000$ ). Most importantly, there was an interaction between trust and label,  $F(1, 398) = 21.99$ ,  $p < .001$ ,  $\eta_p^2 = .052$  (see Figure 6). In the high trust condition, we replicated the basic effect: participants perceived the revision to be significantly more appealing than the original ( $M_{\text{original}} = 4.64$ ,  $SD = 1.54$  vs.  $M_{\text{revised}} = 4.95$ ,  $SD = 1.49$ ;  $F(1, 398) = 11.87$ ,  $p = .001$ ,  $\eta_p^2 = .029$ ). In contrast, in the low trust condition, this robust effect reversed: the revision was deemed significantly *worse* than the original ( $M_{\text{original}} = 4.38$ ,  $SD = 1.66$  vs.  $M_{\text{revised}} = 4.09$ ,  $SD = 1.67$ ;  $F(1, 398) = 10.16$ ,  $p = .002$ ,  $\eta_p^2 = .025$ ).

In sum, consistent with our account, study 7 indicated that the preference for revision was moderated by the reviser's trustworthiness. In fact, the preference for revision was *reversed* when the reviser was untrustworthy, with participants deeming the revision to be worse than the original in this case.

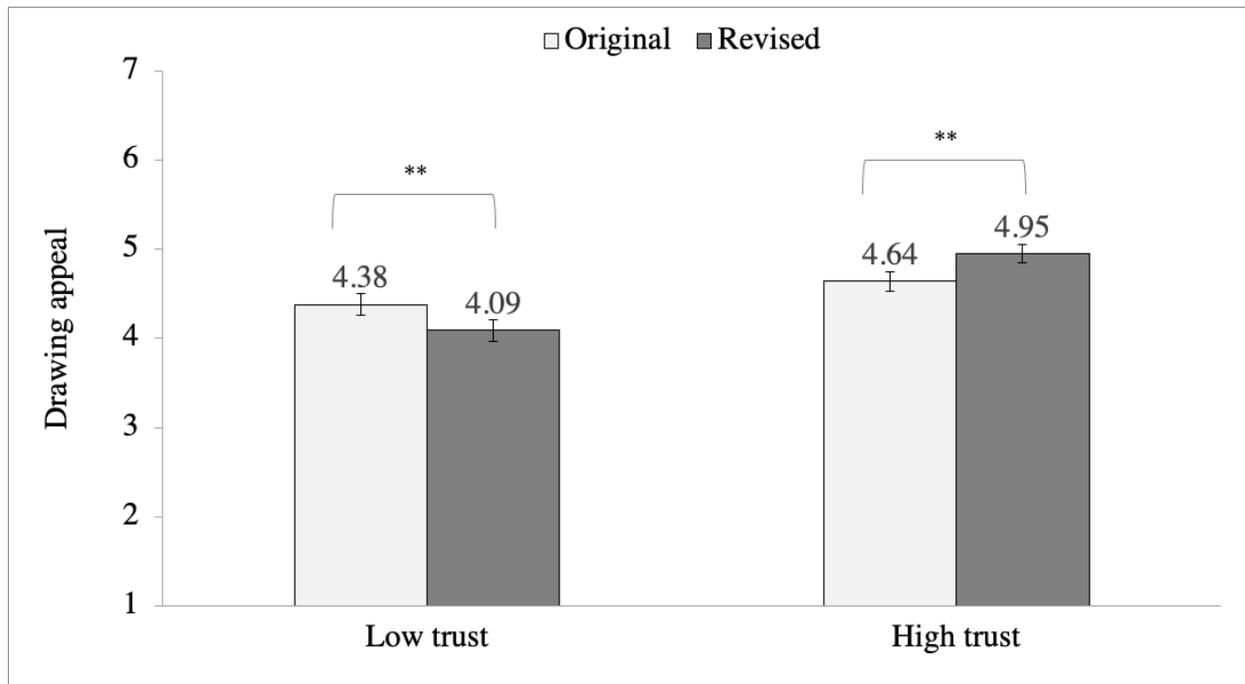


Figure 6. Study 7: mean ratings for original and revised drawing across conditions. Error bars indicate  $\pm 1$  SEM.  $*p < .05$ ,  $**p < .01$ ,  $***p < .001$ .

## GENERAL DISCUSSION

To “revise and resubmit” is far more than academic exercise. Companies often work to revise their products and services, just as individuals often revise their own creations and contributions. In principle, a world of constant revision should lead to better outcomes for both creators and consumers; in practice, the current research counsels caution. Eleven studies document the preference for revision across a wide variety of stimuli and contexts. Our studies

also offer process evidence for when and why this phenomenon emerges. We demonstrate that the preference for revision arises from an overgeneralization of an often-reasonable belief—that revisers are truly trying to improve their creations and that revisions represent the successful fruition of those intentions. In turn, consumers place undue faith in revisions—in a sense, abandoning critical assessment—leading to the perception that revisions are superior to originals, even absent objective improvement.

Studies 1-4 showed the basic effect. In study 1A, MBA students worked to revise their resumes over the course of a semester, but those efforts did not pay off: resumes at the start would have been evaluated just as highly, so long as they were merely labeled as the revised versions (studies 1B and 1C). In study 2, the same candy tasted better merely when participants thought its recipe had been revised; in study 3, participants preferred the same product that had a revised label (over one that had a standard label). In studies 4A and 4B, participants preferred “revised” stimuli even when the alleged revisions were obviously trivial or identical to their original versions. Studies 5-7 document the underlying mechanism. First, studies 5A and 5B demonstrate participants’ belief that revisers both intend to and succeed in improving their creations explain why consumers prefer revised versions, while study 6 shows a consequence of these beliefs: consumers suspend critical assessment, identifying fewer bugs in a video game they believed had been revised. Finally, study 7 shows that the preference for revision is attenuated—and in fact, even reversed—when beliefs about revisers’ good intentions are called into question.

One concern in general with studies of expectancy effects revolves around the role of demand characteristics: participants may be reporting what they believe the experimenter wants them to report. We believe demand is unlikely to account for our full set of studies, in which we

find similar patterns of results across a wide variety of stimuli and contexts. Moreover, specific study features directly address the role of demand. For example, participants in study 3 made a choice without knowing they were part of a study: the study was conducted surreptitiously, during the introductory paperwork that participants completed at the start of a battery of lab studies. Participants were asked to sign an administrative form and could select one of two pens: one pen had a standard label and one pen had a label that said “revised.” Given that in this study the choice was framed as part of an administrative process (rather than as part of a study), the possibility that participants were more likely to choose the revised option due to demand is reduced. Likewise, study 6 uses a fully between-subjects design, and shows that the effect holds when we invoke revision without using the word “revision”—here, we denote revision by version number (e.g. “Version 5.0”).

### Marketing Implications

This preference for revision raises clear marketing implications. On the one hand, products that are objectively unchanged (or even made worse) in the revision process may nonetheless be adopted, so long as consumers *believe* they possess a “revised” version. This may happen innocently (e.g., consumers may be prone to upgrading to a revised version even if that version is not objectively better), but also intentionally (e.g., companies that release annual updates for the sake of releasing annual updates). This preference for revision emerged even after giving participants complete first-hand knowledge of the entity in question (e.g., eating a candy, playing a video game, choosing a pen to sign a form). If consumers cannot easily trust their bottom-up experience to draw more informed conclusions, they could be influenced by companies who are motivated to make revisions for reasons beyond the desire to improve the

quality of their products. On the other hand, things that are objectively improved in the revision process may go unappreciated, if the fact that they have been revised is not made deliberately clear: e.g., a candy's recipe almost certainly underwent extensive editing and tweaking behind the scenes. In sum, the preference for revision raises dual marketing implications for revisions wielding influence when perhaps they should not, and failing to wield influence when perhaps they should, due merely to the label affixed to them.

### Theoretical Implications

This research also makes important theoretical contributions. First, our research extends research on the links between expectations and experience, addressing an issue raised by Lee et al. (2006) in their review of outstanding questions on such links: “A third question concerns how specific perceptual, attentional, and cognitive mechanisms mediate the effect of expectations on experience (or reported experience)” (p. 1057). Extending this research, studying the effects of “revision” labeling explores a novel aspect of the idea that expectations affect experiences: even in the absence of specific information about *how* the experience will be changed (“tastier,” “fattier”), we suggest that mere expectation of change itself (“revised”) influences consumption experiences. Our studies provide finer-grained and novel evidence for how, exactly, expectations change one's experience: in our case, revision labels lead participants to essentially pay less attention and notice fewer flaws in identically buggy stimuli. Providing direct evidence for this specific process—documenting specific behavioral changes—offers insight into the drivers of changes on outcome measures such as liking and preference (as in many of the papers reviewed in Lee et al. 2006).

Second, the preference for revision contributes to understanding an emerging collection of related phenomena that share two features: an ordering or sequencing component and a framing component. For example, building on the tendency for the first item in a given array to be preferred (Carney and Banaji 2012; Murdock and Bennet 1962)—a sequencing effect—research has shown that “phantom firsts”—merely framing something as “first”—increases its appeal (LeBoeuf, Williams, and Brenner 2014). Interestingly, these phenomena suggest that in the absence of explicit revision, stimuli that are framed to have occurred earlier in a sequence are preferred (Smith, Newman, and Dhar 2015).

At the same time, other research points to the notion that items that have been in existence for longer are preferred, as with the longevity bias (Eidelman, Pattershall, and Crandall 2010). While our results cannot be accounted for by the longevity bias—for example, in studies 3, 4B, and 6, the pens, logos, and video games had existed for the same amount of time in both conditions, but framing them as revised versions increased preference for them—clearly more research is needed to understand when different temporal sequences are preferred. Expanding on this point, future research is needed to examine the preference for revision in the context of apparent exceptions, such as when people desire older, “pre-revised,” experiences—as when experiencing nostalgia (Wildschut, Sedikides, Arndt, and Routledge 2006), rediscovering past experiences (O’Brien 2019; Zhang, Kim, Brooks, Gino, and Norton 2014), or valuing original renditions of collectibles over later renditions (e.g., artwork; Newman and Bloom 2012).

Future research could also fruitfully explore other intuitions about revisions. For example, because perceptions that revision has occurred (even when it has not) are critical for influencing perceived quality, it would be valuable to uncover other factors that influence perceived revision, such as content knowledge (e.g., experts may be less susceptible to this

preference for revised things) and the nature of the change (e.g., adding new features may seem like a more substantive revision than deleting existing features: Agostinelli, Sherman, Fazio, and Hearst 1986).

In sum, the preference for revision can in some situations lead consumers astray if the revision was done for the sake of revising things. On the positive side, any author who has experienced the pain of making endless revisions to a manuscript should take heart: getting readers to appreciate one's genuine efforts to improve may prove easier than it seems.

## REFERENCES

- Agostinelli, Gina, Setevn J. Sherman, Russell H. Fazio and Eliot S. Hearst (1986), "Detecting and identifying change: Additions versus deletions," *Journal of Experimental Psychology: Human Perception and Performance*, 12 (4), 445-454.
- Antonovsky, Aaron (1987), *Unraveling the mystery of health: How people manage stress and stay well*, Jossey-bass.
- Baron, Jonathan (1990), "Harmful heuristics and the improvement of thinking," In *Developmental Perspectives on Teaching and Learning Thinking Skills*, vol. 21, pp. 28-47. Karger Publishers.
- Bornstein, Robert F. (1989), "Exposure and affect: overview and meta-analysis of research, 1968–1987," *Psychological Bulletin*, 106 (2), 265-289.
- Buell, Ryan W., and Michael I. Norton (2011), "The labor illusion: How operational transparency increases perceived value," *Management Science*, 57 (9), 1564-1579.
- Bulow, Jeremy (1986), "An economic theory of planned obsolescence," *The Quarterly Journal of Economics*, 101 (4), 729-749.
- Carney, Dana. R., and Mahzarin R. Banaji (2012), "First is best," *PloS ONE*, 7 (6), e35088.
- Chinander, Karen R., and Maurice E. Schweitzer (2003), "The input bias: The misuse of input information in judgments of outcomes," *Organizational Behavior and Human Decision Processes*, 91 (2), 243-253.
- Deliza, Rosires, and Halliday JH MacFie (1996), "The generation of sensory expectation by external cues and its effect on sensory perception and hedonic ratings: a review," *Journal of Sensory Studies*, 11 (2), 103-128.

- Diehl, Kristin, Erica Van Herpen, and Cait Lamberton (2015), "Organizing products with complements versus substitutes: effects on store preferences as a function of effort and assortment perceptions," *Journal of Retailing*, 91 (1), 1-18.
- Eidelman, Scott, Jennifer Pattershall, and Christian S. Crandall (2010), "Longer is better," *Journal of Experimental Social Psychology*, 46 (6), 993-998.
- Epstein, Seymour (1991), "The self-concept, the traumatic neurosis, and the structure of personality," In D. Ozer, J. N. Healy, & A. J. Stewart (Eds.), *Perspectives in Personality* (pp. 63–98). London, England: Kingsley.
- Fishbein, Martin and Icek Ajzen (1976), "Misconceptions about the Fishbein model: Reflections on a study by Songer-Nocks," *Journal of Experimental Social Psychology*, 12 (6), 579-584.
- Fishbein, Martin (1980), "A theory of reasoned action: some applications and implications," In *Nebraska Symposium on Motivation. Nebraska Symposium on Motivation* (Vol. 27, p. 65).
- Gershoff, Andrew D., Ran Kivetz, and Anat Keinan (2012), "Consumer response to versioning: How brands' production methods affect perceptions of unfairness," *Journal of Consumer Research*, 39 (2), 382-398.
- Henard, David H., and David M. Szymanski (2001), "Why some new products are more successful than others," *Journal of Marketing Research*, 38 (3), 362-375.
- Hoch, Stephen J., and Young-Won Ha (1986), "Consumer learning: Advertising and the ambiguity of product experience," *Journal of Consumer Research*, 13 (2), 221-233.

- Hsee, Christopher K. (1996), "The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives," *Organizational Behavior and Human Decision Processes*, 67 (3), 247-257.
- Hsee, Christopher K., George F. Loewenstein, Sally Blount, and Max H. Bazerman (1999), "Preference reversals between joint and separate evaluations of options: a review and theoretical analysis," *Psychological Bulletin*, 125 (5), 576-590.
- Hsee, Christopher K., and Jiao Zhang (2010), "General evaluability theory," *Perspectives on Psychological Science*, 5 (4), 343-355.
- Janoff-Bulman, Ronnie, and Irene H. Frieze (1983), "A theoretical perspective for understanding reactions to victimization," *Journal of Social Issues*, 39 (2), 1-17.
- Janoff-Bulman, Ronnie, and Cynthia McPherson Frantz (1997), "The impact of trauma on meaning: From meaningless world to meaningful life," In M. J. Power & C. R. Brewin (Eds.), *The transformation of meaning in psychological therapies: Integrating theory and practice* (pp. 91–106). Hoboken, NJ: Wiley.
- Klein, Nadav, and Ed O'Brien (2017), "The power and limits of personal change: When a bad past does (and does not) inspire in the present," *Journal of Personality and Social Psychology*, 113 (2), 210-229.
- Koehler, Derek J., Rebecca J. White, and Leslie K. John (2011), "Good intentions, optimistic self-predictions, and missed opportunities," *Social Psychological and Personality Science*, 2 (1), 90-96.
- Kruger, Justin., Derrick Wirtz, Leaf Van Boven, and T. William Altermatt (2004), "The effort heuristic," *Journal of Experimental Social Psychology*, 40 (1), 91-98.

- LeBoeuf, Robyn A., Eleanor F. Williams, and Lyle A. Brenner (2014), "Forceful phantom firsts: Framing experiences as firsts amplifies their influence on judgment," *Journal of Marketing Research*, 51 (4), 420-432.
- Lee, Leonard, Shane Frederick, and Dan Ariely (2006), "Try it, you'll like it: The influence of expectation, consumption, and revelation on preferences for beer," *Psychological Science*, 17 (12), 1054-1058.
- Lerner, Melvin J. (1980), "The belief in a just world," In *The Belief in a Just World* (pp. 9-30). Springer, Boston, MA.
- Loewenstein, George F., and Dražen Prelec (1993), "Preferences for sequences of outcomes," *Psychological Review*, 100 (1), 91-108.
- Mischel, Walter, and Carolyn C. Morf (2003), "The self as a psycho-social dynamic processing system: A meta- perspective on a century of the self in psychology," In M. R. Leary & J. P. Tangney (Eds.), *Handbook of Self and Identity* (pp. 15–43). New York, NY: Guilford.
- Montoya, Amanda K. (2019), "Moderation analysis in two-instance repeated measures designs: Probing methods and multiple moderator models," *Behavior Research Methods*, 51 (1), 61-82.
- Moore, Don A., and Deborah A. Small (2007), "Error and bias in comparative judgment: on being both better and worse than we think we are," *Journal of Personality and Social Psychology*, 92 (6), 972-989.
- Morales, Andrea C. (2005), "Giving firms an "E" for effort: Consumer responses to high-effort firms," *Journal of Consumer Research*, 31 (4), 806-812.
- Murdock, Jr, and B. Bennet (1962), "The serial position effect of free recall," *Journal of Experimental Psychology*, 64 (5), 482-488.

- Newman, George E., and Paul Bloom (2012), "Art and authenticity: The importance of originals in judgments of value," *Journal of Experimental Psychology: General*, 141 (3), 558-569.
- O'Brien, Ed (2019), "Enjoy it again: Repeat experiences are less repetitive than people think," *Journal of Personality and Social Psychology*, 116 (4), 519-540.
- O'Brien, Ed and Michael Kardas (2016), "The implicit meaning of (my) change," *Journal of Personality and Social Psychology*, 111 (6), 882-894.
- Park, Crystal L. (2010), "Making sense of the meaning literature: An integrative review of meaning making and its effects on adjustment to stressful life events," *Psychological Bulletin*, 136 (2), 257-301.
- Petty, Richard E., and John T. Cacioppo (1986), "The elaboration likelihood model of persuasion," In *Communication and Persuasion* (pp. 1-24). Springer, New York, NY.
- Ratner, Rebecca K., and Barbara E. Kahn (2002), "The impact of private versus public consumption on variety-seeking behavior," *Journal of Consumer Research*, 29 (2), 246-257.
- Smith, Rosanna K., George E. Newman, and Ravi Dhar (2015), "Closer to the creator: Temporal contagion explains the preference for earlier serial numbers," *Journal of Consumer Research*, 42 (5), 653-668.
- Taylor, Shelley E., and Jonathan D. Brown (1988), "Illusion and well-being: a social psychological perspective on mental health," *Psychological Bulletin*, 103 (2), 193-210.
- Taylor, Shelley E., and Jonathan D. Brown (1994), "Positive illusions and well-being revisited: separating fact from fiction," *Psychological Bulletin*, 116 (1), 21-27.
- TIME (2017), "Jimmy Kimmel pranks clueless people into thinking the iPhone 4 is the iPhone X," In B. Perrigo (Ed.), *TIME*.

- Tormala, Zachary L., Jason S. Jia, and Michael I. Norton (2012), "The preference for potential," *Journal of Personality and Social Psychology*, 103 (4), 567-583.
- Wildschut, Tim, Constantine Sedikides, Jamie Arndt, and Clay Routledge (2006), "Nostalgia: Content, triggers, functions," *Journal of Personality and Social Psychology*, 91 (5), 975-993.
- Wilson, Anne E., and Michael Ross (2001), "From chump to champ: People's appraisals of their earlier and present selves," *Journal of Personality and Social Psychology*, 80 (4), 572-584.
- Wortman, Camille B. (1983), "Coping with victimization: Conclusions and implications for future research," *Journal of Social Issues*, 39 (2), 195-221.
- Wortman, Camille B., and Roxane Cohen Silver (1992), "Reconsidering assumptions about coping with loss: An overview of current research," In L. Montada, S. Filipp, & M. J. Lerner (Eds.), *Life crises and experiences of loss in adulthood* (pp. 341-365). Hillsdale, NJ: Erlbaum.
- Zhang, Ting, Tami Kim, Alison W. Brooks, Francesca Gino, and Michael I. Norton (2014), "A "present" for the future: The unexpected value of rediscovery," *Psychological Science*, 25 (10), 1851-1860.
- Zajonc, Robert B. (1968), "Attitudinal effects of mere exposure," *Journal of Personality and Social Psychology*, 9 (2p2), 1-27.