HARVARD | BUSINESS | SCHOOL

The Revision Bias

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

Working Paper 19-087

The Revision Bias

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

The Revision Bias

Ximena Garcia-Rada[1], Leslie K. John[1], Ed O'Brien[2], Michael I. Norton[1]
[1]Harvard Business School
[2]Univeristy of Chicago Booth School of Business

Author Note

Ximena Garcia-Rada, Marketing, Harvard Business School; Leslie K. John, Negotiations, Organizations and Markets, Harvard Business School; Ed O'Brien, Behavioral Science, University of Chicago Booth School of Business; Michael I. Norton, Marketing, Harvard Business School.

We thank Derek Koehler for helpful feedback and Holly Howe, Trevor Spelman, and Shun Wang for research assistance. To collect data for Study 1, we worked closely with the team in a professional development office at a business school and are grateful for their support.

Abstract

Things change. Things also *get* changed—often. Why? The obvious reason is that revising things makes them better. In the current research, we document a less obvious reason: Revising things makes people *think* they are better, absent objective improvement. We refer to this phenomenon as the revision bias. Nine studies document this effect and provide insight into its psychological underpinnings. In Study 1, MBA students perceived their revised resumes to be of higher quality the more they differed from their original versions, but this perception was not justified: observers judged originals (inaccurately) labeled as revisions to be superior to revisions (inaccurately) labeled as originals. Study 2 pinpoints the direction of the effect: revisions are appealing, as opposed to originals being unappealing. Moreover, the revision bias holds in a variety of settings in which the revision is devoid of objective improvement: when revisions are trivial (Study 3A), incidental (Study 3B), non-existent (Study 3C), and even objectively worse than the original (Study 3D). Study 4 directly tests the self-fulfilling nature of the revision bias, testing whether mere revision framing leads people to become less critical of the experience—in this study, less sensitive to possible bugs while playing an otherwise identical "revised" video game—and whether this mediates the effect of revision framing on positive evaluations. Studies 5A and 5B offer further support by testing whether the revision bias is accentuated when people engage in a holistic processing style, whether measured as an individual difference (Study 5A) or experimentally induced (Study 5B).

Keywords: change, heuristics and biases, framing, sequences, judgment

The Revision Bias

Things change. Things also *get* changed—often. Companies regularly release revised editions of books, director's cuts of movies, and technological updates. Individuals regularly tweak recipes, edit resumes, and revise manuscripts. Why? The obvious reason is that revising things makes them better. In the current research, we document a less obvious reason: Revising things makes people *think* they are better, absent objective improvement, even when people can check for themselves by seeing and using those unimproved things before forming a judgment. We refer to this phenomenon as the revision bias.

The revision bias is distinct from general preferences for novelty (Berlyne, 1970; Sheldon, Boehm, & Lyubomirsky, 2012), and from general preferences for recency (O'Brien & Ellsworth, 2012), as revisions can be compared to unrevised counterparts from the same points in time. Instead, we propose that the fact that revisions typically are intended to be improvements over their originals gives rise to an overgeneralized heuristic that revisions necessarily *are* improvements over their originals. Yet, as any author responding to editorial reviews knows, not every revision turns out better than before. The revision bias, like other heuristic-based judgment processes, may be valid in many cases but unwittingly applied in other cases even when it does not reflect reality (Kahneman & Frederick, 2002; Kahneman & Tversky, 1973; Simon, 1957; Tversky & Kahneman, 1975).

We propose that the revision bias is self-fulfilling, wielding a top-down influence that alters how observers experience the stimulus for the first time. As people test out a "revised" gadget, for example, they may be less likely to search for or recognize potential bugs merely by virtue of its label, and thus come away concluding that the "revised" gadget is as great as advertised. In a memorable public demonstration (see TIME, 2017) around the release of

Apple's iPhone X (November 2017), city passersby were invited for a sneak-peak to play with the phone. Unsurprisingly, many of them raved about the phone's sleek new feel and performance. In reality, all of them had been handed a well-worn iPhone 4 (released June 2010). This possibility of a top-down effect of mere revision labels, in the absence of objective improvement, is especially problematic given that revisions can be strategic, such as when companies release annual updates so as to renew public attention and not because change is warranted ("If it ain't broke, fix it anyway"). If observers cannot trust their own consumption experience to assess the quality of a revision, the revision bias may lead people astray.

Why do we hypothesize the existence of a revision bias? Heuristics give rise to systematic errors in judgment, with two relevant tendencies. First, formally equivalent ways of describing a stimulus produces different judgments (Sher & McKenzie, 2011); for example, people are more accepting of a program that incurs substantial loss of life when described in terms of lives lost versus lives saved (Tversky & Kahneman, 1981). Second, judgment is inherently comparative which, combined with a reliance on heuristics, makes people susceptible to arbitrary reference points, or anchors (Furnham & Boo, 2011; Hsee & Zhang, 2010; Mochon & Frederick, 2013); for example, people's willingness to pay for products can be influenced by irrelevant reference prices (Ariely, Loewenstein, & Prelec, 2003; Tversky & Kahneman, 1975), while their effort can be influenced by arbitrary reference points (Barasz, John, Keenan, & Norton, 2017; Heath, Larrick, & Wu, 1999). These tendencies are amplified when targets of judgment are ambiguous, leading people to seek help from contextual cues like frames and anchors, which can then influence their subsequent experiences of the target (Bless & Schwarz, 2010; Higgins, 1996; Klaaren, Hodges, & Wilson, 1994; Lee, Frederick, & Ariely, 2006).

Susceptibility to arbitrary frames and anchors forms the basis of our hypothesis. Because judgment is inherently comparative, encountering a revised or new version prompts a natural comparison group: an earlier version. This comparison is likely to be top of mind when a "revised" version of a stimulus is juxtaposed against an original version. In turn, because judging quality can be ambiguous (in particular when encountering a stimulus for the first time), people may be influenced by the label even if it is irrelevant, uninformative, or meaningless in reality—even still if people get first-hand experience with the stimulus themselves, as their experience may be tainted by the label.

Further supporting our hypothesis that revisions will specifically elicit *positive* evaluations, other psychological research documents the high value that people place on change, often for sake of change, all else equal (Furnham, 1984; Hsee & Abelson, 1991; Loewenstein & Prelec, 1993; Ryff, 1991; Wilson & Ross, 2001). Looking forward, for example, people express more interest in others who have high potential to improve than in others who are currently at that same positive state (Tormala, Jia, & Norton, 2012). Looking back, people feel more inspired by others who have changed from exhibiting bad-to-good character over time than by others who have worked to maintain that same good character all along (Klein & O'Brien, 2017). Most relevant to our hypothesis, O'Brien and Kardas (2016) documented an automatic association between the concept of change and *positive* reactions: participants who were asked to reflect on how they had changed over time, with no definition of "change," brought to mind only ways they had *improved*—despite easily bringing to mind decline when specifically prompted. Thus, the mere presentation of a "revised" stimulus, with no additional information or evidence of improvement, nonetheless may lead people to view the stimulus in a distinctly favorable light.

Finally, because assessing whether or not a revision actually represents an improvement over an original requires attention to detail in a more bottom-up manner (as opposed to relying on a top-down heuristic that revisions are always better), the revision bias, like other biases, is likely undergirded by an intuitive, or "System 1" processing style (Kahneman, 2011; Kahneman & Frederick, 2002). Thus, in addition to the top-down effect of revision labels as described earlier, we propose that the revision bias should be more likely to emerge under conditions that specifically draw attention to the "big picture" rather than to smaller details that may actually contribute to stimulus quality.

Previous research has documented reliable differences in people's tendency to think holistically (fast, global) versus analytically (deliberative, effortful; Chaiken, 1980; Chaiken & Maheswaran, 1994). According to our account, holistic thinkers may be more prone to the revision bias, given that their thinking style dictates that they are less attentive to detailed comparisons relative to analytical thinkers. We propose that the revision bias is particularly likely to emerge when people process stimuli holistically rather than analytically, whether due to an individual difference or even experimentally induced. In sum, people generally may be influenced by mere revision framings, sticking to the surface and evaluating a "revised" stimulus in a passively positive light; by the same logic, however, knowing to engage with a "revised" stimulus in more analytical ways may serve to combat the revision bias.

Overview of Studies

Nine studies document the revision bias and provide insight into its psychological underpinnings. Study 1 uses field data to test for the revision bias among MBA students engaged in a resume revision process. Study 2 then replicates the revision bias in the laboratory using an experiential stimulus (eating gummy candy), and pinpoints its direction: revisions are appealing,

as opposed to originals being unappealing. Next, in more direct support of our contention that the revision bias is in fact a bias—a judgment error—we test whether the revision bias holds when it is transparent that the revision is trivial (i.e., does not represent improvement over the original, Study 3A), incidental (Study 3B), and non-existent (Study 3C), potentially even leading people to prefer revised stimuli over objectively superior (but unrevised) stimuli (Study 3D). Finally, Studies 4, 5A, and 5B assess whether the revision bias operates as a judgment heuristic as proposed. Study 4 directly tests the self-fulfilling nature of the revision bias, testing whether mere revision framing leads people to become less critical of the experience—in this study, less sensitive to possible bugs while playing an otherwise identical "revised" video game—and whether this mediates the effect of revision framing on positive evaluations. Studies 5A and 5B offer further support by testing whether the revision bias is accentuated when people engage in a holistic processing style, whether measured as an individual difference (Study 5A) or experimentally induced (Study 5B).

For online studies, we pre-specified the sample sizes to 75 to 250 per cell, depending on study design. For Study 1, we collected data from all students who agreed to participate. For Study 2, we collected as much data as we could in three days of laboratory sessions. Study 4 was preregistered at http://aspredicted.org/blind.php?x=r42sj2. We report all manipulations and measures and did not analyze data until collection was complete. Data and stimuli for all studies except Study 1 are at: https://osf.io/4x76c/?view_only=b3e050d5cb774780ac2a576b84660eca (we did not post data from Study 1 because we do not have permission to do so from the university office with which we collaborated). Unless otherwise indicated, no data were excluded and attrition rates were low (statistically equivalent across conditions and never exceeding 5%).

Study 1: Field Study

Study 1 was a two-phase field study conducted during a process designed to help MBA students of a Northeastern US business school prepare their resumes. MBA students submit an initial version of their resume, are given opportunities to revise it based on feedback from resume coaches, and submit a final version for inclusion in a "resume book" distributed to prospective employers. In Phase 1, 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 Phase 2, we tested whether this perception is warranted: observers rated two versions of an MBA student's resume. Half of observers simply rated the actual original, followed by its corresponding actual revision. For the other half of observers, we swapped the labels put on each version: 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). Here we predicted that observers would show no preference for resumes that were actually revised—suggesting that revised versions are not objectively better—but would rate any versions labeled "revised" to be better than those labeled "original," demonstrating the revision bias.

Methods

Phase 1: Stimuli development. In Phase 1, we collaborated with the business school's Career and Professional Development (CPD) team, following their process: students submit an initial resume version to CPD; over two months, students can receive feedback from a resume coach; students submit a final resume for inclusion in a resume book distributed to prospective employers. This granted us the opportunity to analyze a diverse set of genuinely revised stimuli.

In the fall of 2016, 302 students (demographic information not assessed) 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% (233 out of 302) of students who agreed. We sent an email to these 233 students, hereafter referred to as "authors," containing a link to a short survey (see appendix A in Supplementary Procedure 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 authors 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 authors indeed revised their resumes and viewed the final products as improved—presumably their goal in having formally worked on their resumes over the course of the semester.

At the end of the survey, authors were informed about Phase 2 and could opt out of letting us use their resumes. In total, 6 authors failed to answer at least one of these survey questions, and an additional 5 authors opted out. This left a final $N = 33$ authors (i.e., 33 pairs of resumes) to be used for Phase 2.

Phase 2: Main study. In Phase 2, an independent sample of participants, hereafter

referred to as "observers," evaluated the resumes. Observers (N = 453, 50.1% male; M_{age} = 30.27 years, SD = 12.27) were community members who came to the laboratory to participate in this and a series of unrelated studies. The sessions were run between December 2016 and January 2017. Because the unrelated studies in the session were different in December versus January, it was possible for people to participate both months. We identified individuals who participated in both sessions using lab identifiers. We report the results using the full sample; however, our results hold when excluding the (January) data from those indicating that they had already participated (see appendix A in Supplementary Procedure).

Observers rated both versions of one resume pair (i.e., one MBA student's original version and the same MBA student's revised version), shown on a computer screen. The name of each author was replaced with the generic name "Alex Newman" with all other author identifiers removed. 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

Revision validation (author ratings). Authors deemed their revised resume to be significantly different from their original resume (M = 36.34%, SD = 26.33; 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; t-test against the scale midpoint, $t(43) = 11.38$, $p < .001$). Moreover, the more dissimilar their two versions, the higher quality the authors perceived their revised version, $r = .612$, $p < .001$; $N = 38$[1]. This relationship held, $\beta = .58$, SE = .004, $t(35) = 4.88$, $p < .001$, over and above the effect of the number of times the author had obtained feedback (which independently predicted perceptions of improvement, $\beta = .37$, SE = .113; $t(35) = 3.07$, $p = .004$). Authors 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 authors' eyes, should be noticed by others as significantly improved.

The revision bias (observer ratings). Authors were mistaken: observers exhibited the revision bias (see Figure 1). A 2 × 2 mixed ANOVA revealed that while actual revised resumes

[1] The valid sample for this specific test was 38 subjects. This is because even though 44 authors completed the questionnaire of Phase 1, six of them left blank the question "What percentage of the final version of your resume is different from your original resume?".

were seen as no better than actual original resumes, $F(1, 451) = 0.098$, $p = .754$, $\eta_p^2 = .000$, 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, $t(226) = -4.20$, $p < .001$, $d = 0.293$. In other words, authors' efforts only mattered if observers were told the resume was revised; all of those edits they made to their resumes did not, in and of themselves, improve observers' assessments.

As further evidence for the revision bias, we compared authors' ratings of how different their revisions were from their originals with observers' ratings of those revisions; these were uncorrelated ($r = .047$, $p = .318$), suggesting that revision—the extent to which resumes changed according to their authors—did not predict observers' evaluations; instead, the *label* "revision" drove observers' perceptions. In short, it is not attention to the details that have changed that lead to perceptions of improvement, but a holistic evaluation that revisions are better.

Study 1 documents the revision bias in a field setting. The more that students perceived their resumes to have changed, the higher quality they deemed their final versions to be—but this perception was not validated by observers. Instead, observers showed a preference not for truly revised resumes, but for resumes merely labeled as revised (absent any objective improvement).

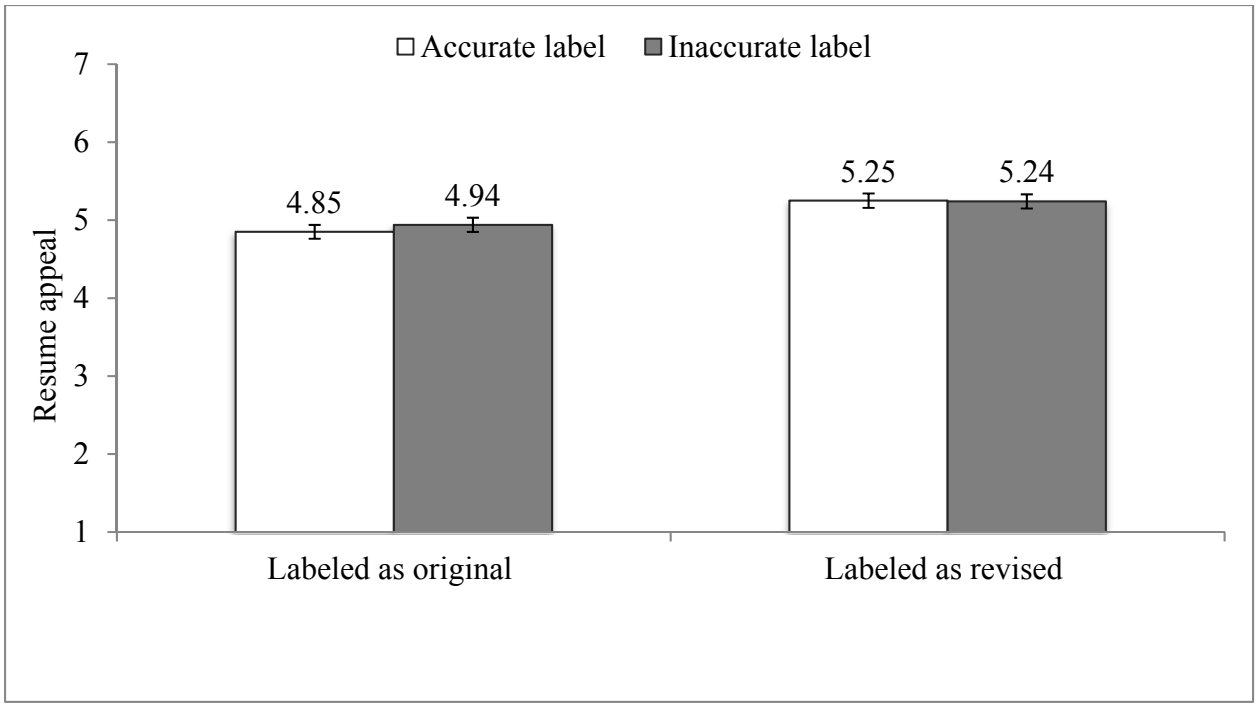

Figure 1. Study 1, Phase 2: mean ratings of resumes across conditions. Error bars indicate ±1 SEM.

Study 2: Specificity of the Effect

Study 2 has three primary goals. First, it pinpoints the direction of the effect, disambiguating that it is revisions that are appealing and not originals that are unappealing, by removing the label "original" from the control stimulus. Second, it tests whether the revision bias generalizes to an experiential stimulus: eating gummies. Third, given the tendency to reward effort (Buell & Norton, 2011; Chinander & Schweitzer, 2003; Kruger, Wirtz, Van Boven, & Altermatt, 2004; Morales, 2005), Study 2 also assesses whether the revision bias emerges over and above any increase in value due to perceived effort.

Methods

Participants (N = 239; 49.8% male; M_{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 appendix B in Supplementary Procedure).

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*).

Next, participants were told that the remaining gummy—the gummy on the right, which served as the experimental gummy—had been revised and that "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. Note that the control stimulus (in this case, gummy) was given no special label—i.e., it was not labeled "original" as in Study 1.

Results

Again, participants exhibited the revision bias. A composite measure of the two primary outcome measures (quality and taste, $\alpha_{control} = .83$, $\alpha_{revised} = .88$) revealed that the revised gummy was deemed superior to the control gummy ($M_{revised} = 3.77$, $SD = 1.77$; $M_{control} = 3.47$, $SD = 1.57$), $t(238) = -2.41$, $p = .017$, $d = 0.16$—despite the fact that the gummies were *merely labeled*

this way (effects hold when the measures are analyzed separately, see appendix B in Supplementary Procedure). 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 ($M_{revised} = 3.99$, $SD = 1.80$; $M_{control} = 3.41$, $SD = 1.63$), $t(238) = -4.94$, $p < .001$, $d = 0.32$). The results are also substantively equivalent when excluding the 22 participants who reported that they did not sample both gummies (see appendix B in Supplementary Procedure for robustness checks).

Study 2 replicates the revision bias in the laboratory, using a stimulus that gave participants the entire bottom-up experience (i.e., actually eating the gummies) before making their evaluations. Moreover, the results show that the revision bias is driven by the unique appeal of revisions (as opposed to originals being unappealing), and emerges over and above perceptions of increased effort.

Studies 3A–3D: The Bias in Revision Bias

Studies 3A–D explore whether the revision bias is indeed a judgment bias, testing whether the preference for revision emerges when the revision is obviously trivial (Study 3A) and when the revision label is obviously incidental (Study 3B). Studies 3B and 3C further address whether the effect can be accounted for by reasonable inferences—for example, that the creator knows something about appeal that the participant does not (Studies 3B and 3C). Finally, Study 3D tests whether labeling a stimulus as "revised" can even cause people to prefer it to an objectively superior counterpart.

Study 3A: Methods

Study 3A tests whether the revision bias holds when the revision represents a trivial change: a resume is "revised" to a different—but similar—font, with the content unchanged. In addition, Study 3A assessed the revision bias at a perceptual level, examining whether the attributes of a font (e.g., readability, clarity) are perceived differently as a function of revision framing.

Participants (N = 401; 50.6% male; M_{age} = 37.40 years, SD = 12.10) from Amazon's MTurk completed the 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) and presented on the same page (see appendix C in Supplementary Procedure for exact resumes). 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 1, 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*.

Study 3A: Results

A 2 × 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 2A). The revision bias again emerged, 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 2A) 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 the revision bias. Holding font constant, a given font was perceived as better on all measured attributes—clarity, professionality, readability, niceness, and appeal—when it was used in the version labeled "revised" relative to the version labeled "original" (see Figure 2B); 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 $ps < .001$).

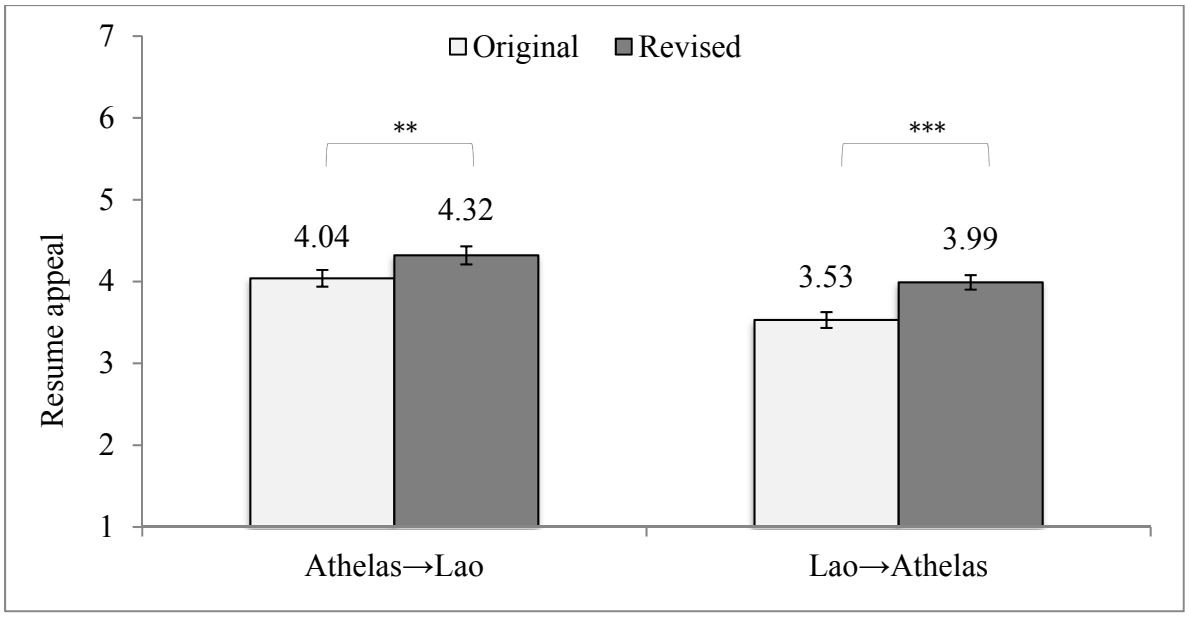

Figure 2A. Study 3A: mean ratings for original and revised versions across conditions. Error bars indicate ±1 SEM. Significance levels compare the two versions in each condition (**$p < .01$, *** $p < .001$).

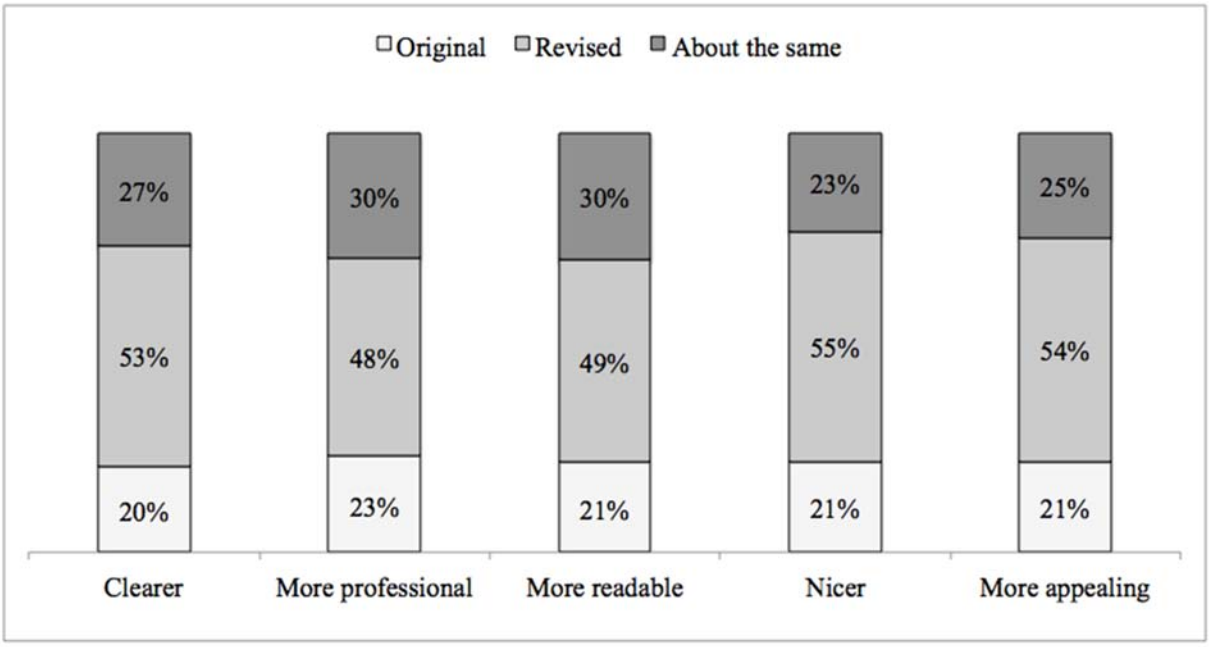

*Figure 2B***.** Study 3A: participants' assessments of which font was superior for each of five attributes. The figure plots, for each attribute, the proportion of participants indicating that the font on the original version was superior (white above); the font on the revised version was superior (light grey above); or that the neither font was better than the other (dark grey above).

In sum, Study 3A shows that the revision bias holds when the revision is obviously trivial, and this triviality was made salient to participants by presenting both resumes on the same page. Study 3A also suggests that the revision bias affects people at a perceptual level; the fonts were perceived to have different attributes as a function of the (arbitrary) revision label.

One potential concern with Study 3A is that participants may have inferred that the author believed the revised resume was superior (despite appearing obviously trivial to them), or else the author would not have chosen to name that file "revised." Study 3B rules out this possibility by making the revision incidental, and explicitly conveying this fact to participants.

Study 3B: Methods

Participants (N = 312; 49.0% male; M_{age} = 33.68 years, SD = 10.13) from Amazon's MTurk completed the study, which was the second of four unrelated studies administered together for a fixed payment. Participants rated the appeal of each of two versions of a resume differing only in font and, as in Study 3A, between-subjects, we manipulated the ascription of revision label to resume.

Participants read: "A person goes to edit their resume. They open up the original version on their word processor. Here is the original draft." Directly below was a screen shot of a resume; below, participants rated its appeal using the same scale as Study 3A. On the next page, participants were told: "the person is using a word processor and this program auto-saves every 2 minutes; every 2 minutes it appends a number to the filename representing the version number." Directly below was a screen shot of a resume that differed only in font; below, participants rated its appeal. This design rules out incidental inferences about the authors' intentions, isolating the effect of mere revision labeling.

On the next page, participants answered two comprehension check questions, one requiring them to identify that they had been shown two versions of the same person's resume (94.2% passed); the other testing correct identification that the person "used a word processor that auto-saved the document every 2 minutes" amid five distractor options (82.4% passed).

Study 3B: Results

The revision bias still emerged. A 2 × 2 mixed ANOVA on overall appeal revealed only the critical main effect of revision label: whatever resume was merely labeled "revised" (M = 3.91, SD = 1.25) was perceived as more appealing than the one labeled "original" (M = 3.44, SD = 1.24), $F(1, 310) = 95.35$, $p < .001$, $\eta_p^2 = .235$. There was no main effect of ascription of font to resume, $F(1, 310) = 0.24$, $p > .25$, $\eta_p^2 = .001$, nor was there an interaction, $F(1, 310) = 0.35$, $p > .25$, $\eta_p^2 = .001$. The revision bias holds when restricting the sample to those passing the comprehension checks (see appendix D in Supplementary Procedure).

Studies 3A and 3B show the revision bias holds when revisions are obviously trivial and when revision labels are obviously incidental. Nonetheless, it remains possible that participants could have made other inferences that bear on the effect, for example that the change was spurred by the author's discovery of employer font preferences. Study 3C rules out such possibilities by documenting the revision bias in a set-up in which the "revision" is literally identical to the original.

Study 3C: Methods

Participants (N = 409; 59.4% male; M_{age} = 34.48 years, SD = 10.47) from Amazon's MTurk completed the study, which was the first of three unrelated studies administered together for a fixed payment. Participants rated the appeal of two different versions of a logo for "Cleansy" (a fictitious soap) across three presentations, using the same scale as used in Studies

3A and 3B. Participants first saw and rated the appeal of one of the logos, labeled "original version." This logo served as the focal logo, because it was 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 "2nd revised version" (see Table 1 for the logos and overview of the design). To ensure the arbitrariness of revision status, between-subjects, we manipulated the ascription of logo to focal logo status. At the end of the study, participants answered a comprehension check question assessing their knowledge of the number of logos presented (see appendix E in Supplementary Procedure).

Our primary prediction was that the 3rd-presented logo would be more appealing when labeled "2nd revised logo" relative to when labeled "back to the original version." Because this contrast is restricted to the 3rd-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 1st-presented logo). Note that this design rules out any incidental inferences about intent (e.g., in both cases, it should seem equally likely that the firm had discovered some font preference that prompted the return to the first design). Furthermore, we also predicted that the first-presented logo (labeled "original version") would be less appealing than the second-presented logo (a different logo, labeled "revised version"), a conceptual replication of the basic effect in all other studies so far.

Table 1.

Overview of design

	1st-presented logo (focal logo)	2nd-presented logo (a different logo)	3rd-presented logo (focal logo again)
			
Control	"original version"	"revised version"	"back to the original version"
Experimental	"original version"	"revised version"	"2nd revised version"

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

Study 3C: Results

A comparison of the appeal of the 3rd-presented logo between conditions tests our primary prediction. A 2 × 2 between-subjects ANOVA controlling for logo order revealed a main effect of revision label: the identical logo was more appealing when it was labeled "2nd revised version" (M = 3.54, SD = 1.46) relative to when labeled "back to the original version" (M = 3.26, SD = 1.43), $F(1, 405) = 3.91$, $p = .049$, $\eta_p^2 = .010$). Thus, the revision bias held even when a revision was literally identical to an original version, from an identical point in time, but merely labeled "revision." The revision bias also 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, $F(1, 405) = 0.46$, $p > .25$, $\eta_p^2 = .001$.

As a conceptual replication of the basic effect in all other studies so far, we compared the appeal of the 1st versus 2nd-presented logo. A 2 × 2 mixed ANOVA revealed a main effect of revision label, $F(1, 407) = 36.08$, $p < .001$, $\eta_p^2 = .081$: the 2nd-presented logo, which was labeled "revised version" (M = 3.42, SD = 1.36) was more appealing relative to the 1st-presented logo,

which was labeled "original version" (M = 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, $F(1, 407) = 0.32$, $p > .25$, $\eta_p^2 = .001$.

Next, Study 3D tests whether framing a stimulus as "revised" might even increase people's preference for it over an objectively superior (but unrevised) stimulus.

Study 3D: Methods

Participants (N = 400; 52.5% male; M_{age} = 34.52 years, SD = 11.05) from Amazon's MTurk completed this study for a fixed payment. Participants were instructed to imagine that they wanted to buy a selfie stick, and that they had narrowed their search to two options. They were given a chart comparing the two options, including a product photo, dimensions, and features (see appendix F in Supplementary Procedure). The only substantive difference between the two selfie sticks was their extension length: one extended to 16 inches and the other extended to 24 inches (the objectively superior option). Objective superiority had been confirmed in a pre-test: a separate sample of participants from the same population (N = 403; 56.3% male; M_{age} = 34.41 years, SD = 10.41) rated two selfie-sticks: a 16-inch stick and a longer one (each rated in terms of overall quality on a 7-point scale from "very low" to "very high"). Between-subjects, we manipulated the length of the longer stick to be either 20-inch, 24-inch, 28-inch, or 32-inch. The results of this pre-test confirmed that longer sticks were deemed better, for all conditions (ps < .001).

In the main study, we manipulated revision framing between-subjects. In the experimental condition, an additional piece of information was provided for each selfie stick at the bottom of the product comparison grid. For the 16-inch stick, this text read: "Additional information: this product was revised by the manufacturer. This is the revised version." For the

24-inch stick, the text read "Additional information: This product is the original version created by the manufacturer." The control condition was identical except did not include this additional information. Two comprehension check questions were administered at the end of the survey.

Study 3D: Results

The inferior selfie stick was more likely to be chosen when it was framed as revised, relative to when no such label accompanied it (percent choosing 16-inch stick in: Experimental = 17.3%; Control = 7.9%), $\chi^2(1) = 8.04$, $p = .005$, Cramer's V = .142). Mere revision framing led participants to be swayed by an objectively worse (but "revised") product.

Finally, Studies 4, 5A, and 5B go beyond documenting the revision bias by also directly examining its psychological underpinnings.

Study 4: Top-Down Mediation

Study 4 has two primary goals. First, we test whether the revision bias replicates using the same single experiential stimulus—playing a novel video game—while merely framing the stimulus as more or less revised. Second, we test finer-grained insights into what participants are actually doing while experiencing the stimulus that may foster more positive evaluations. Our framework posits that mere revision framing may operate as a judgment heuristic, and thus lead people to evaluate a stimulus from a more top-down rather than bottom-up perspective. In this context, we test whether participants are less critical of ambiguous "bugs" when they believe the game has been recently revised, mediating the effect of revision labels on boosted enjoyment for the game—in a sense, rendering perceived improvement to be self-fulfilling.

Methods

Participants ($N = 500$; 58.2% male; $M_{age} = 34.16$ years, $SD = 9.59$) from Amazon's MTurk completed the study for a fixed payment. To begin, all participants read that we—the

requesters of the MTurk HIT—had been working on developing a game called *ART Time*, and they read 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 (see appendix G in Supplementary Procedure for the game). We hired a developer to create the game for purposes of 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—nothing that obviously disrupted the game but that may be possible to interpret as an unintended bug if one was inclined (e.g., the cursor would sometimes lag a split-second behind).

Between-subjects, all participants then read that, to date, we had released 5 updates. We displayed cover-art images of the game, labeled differently in the following way: "Version 1.0, Update: 10/08/2015"; "Version 2.0, Update: 03/23/2016"; "Version 3.0, Update: 08/07/2016"; "Version 4.0, Update: 12/11/2016"; and "Version 5.0, Update: 03/28/2017." (the study was conducted on 05/18/2017). No other information was provided. Participants in the experimental condition were informed that they had been randomly assigned to play Version 5.0 (i.e., the most recent "revision"), whereas control participants were informed that they had been randomly assigned to play Version 2.0 (i.e., a past "revision").

Next, all participants played the game for 2 minutes, during which all other keyboard controls were disabled. When time expired, the page automatically continued to a survey screen where participants rated 2 blocks of questions (the order of the blocks, and the items within each block, were randomized). Each item was rated on a 1 (*not at all*) to 7 (*extremely*) scale. In the enjoyment block (dependent variable), participants rated 5 items: 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 buggyness block (mediator), participants rated 5

items: how buggy and high-tech the game felt while playing, and whether they felt like what they had experienced could be improved, felt as good as it could be, and felt outdated (hereafter referred to as the game buggyness scale, reverse coding where appropriate such that higher scores reflect more perceived bugs; $\alpha = .78$).

Most critical for this study, in reality all participants had played the same exact game. That is, all participants were endowed with the same bottom-up information about the game, experienced in full first-hand. Mere revision framing should not bear on evaluating the game's judged quality, because participants can simply rely on the actual experience that they just had. However, we tested whether participants were nonetheless influenced, such that participants who thought they were playing the most recently "revised" version may be less likely to encode the ambiguous bugs as bugs—thereby creating a more enjoyable experience while playing, despite the fact that the game was identical in both conditions and thus was no objectively better.

Results

Three critical findings emerged. First, the revision bias nonetheless again replicated: an independent samples *t*-test on enjoyment found that participants enjoyed their playing experience significantly more when they merely thought they were playing the most recently revised version ($M = 5.56$, $SD = 1.39$) as opposed to an earlier version ($M = 5.17$, $SD = 1.59$), $t(498) = 2.93$, $p = .004$, $d = 0.26$—despite the fact that all participants had actually played the same exact game. Second, we found converse patterns for buggyness: participants reported that they experienced fewer bugs while playing merely when they merely thought they were playing the most recently revised version ($M = 3.64$, $SD = 1.24$) as opposed to an earlier version ($M = 4.43$, $SD = 1.10$), $t(498) = 7.52$, $p < .001$, $d = 0.67$. Third, the effect of revision framing on game enjoyment was

indeed mediated by differences in experienced buggyness while playing (PROCESS Model 4 at 1,000 iterations: Indirect Effect = .62, Boot SE = .09, 95% CI_{boot} [0.44, 0.80]).

Study 4 replicates and extends the basic effect documented in previous studies. Mere revision framing appears to operate as a judgment heuristic, changing how people experience otherwise identical bottom-up information in ways that are consistent with the frame.

Studies 5A–5B: Moderation by Intuitive Processing

Studies 5A–5B further test for whether the revision bias reflects a judgment heuristic, extending Study 4 by directly assessing intuitive processing via a moderation-based approach—both in terms of individual differences (Study 5A) and a situational manipulation (Study 5B).

First, Study 5A returns to direct comparisons of original-versus-"revised" stimuli to test whether the revision bias is moderated by individual differences in analytic-holistic thinking style (Chaiken, 1980; Chaiken & Maheswaran, 1994). Participants rated an original and revised version of a stimulus (with the ascription of revision label to stimulus manipulated between-subjects) and completed a well-established analytic-holistic thinking style scale that reliably assesses individual differences on this dimension (Choi, Koo, & Choi, 2007). Following our account that the revision bias—like other judgment biases—would be most likely to emerge under holistic rather than analytic processing, we predicted that the revision bias would be pronounced among holistic thinkers (who are less likely to conduct a detailed comparison of original versus revised versions).

Study 5A: Methods

Participants (N = 502; 53.8% male; M_{age} = 35.83 years, SD = 10.67) from Amazon's MTurk completed the study, which was the second of three unrelated studies administered together for a fixed payment. Participants read that an architect had drawn two versions of a

building. Participants first saw and rated the appeal of an original version on a 7-point scale, from 1 (*very low*) to 7 (*very high*). On the next page, they saw and rated the appeal of a revised version of the drawing. The difference between these versions was obviously trivial (the trim was of slightly different color; see appendix H in Supplementary Procedure for exact photographs), and the ascription of revision label to drawing was randomized between-subjects. After rating the two designs, all participants completed the Analysis-Holism Scale (Choi et al., 2007), a 24-item inventory ($\alpha = .77$) that assesses stable differences in these thinking styles. A sample item is, "It is important to pay attention to the whole context rather than the details." Each item is rated on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*), with higher scores reflecting a stronger holistic thinking style (i.e., we predict that participants who score highly on this scale should be more prone to the revision bias).

Study 5A: Results

The revision bias again emerged. A 2 × 2 mixed ANOVA revealed the critical main effect of revision label, such that whatever drawing was merely labeled "revised" was perceived as more appealing ($M = 4.72$, $SD = 1.48$) than the one labeled "original" ($M = 4.32$, $SD = 1.46$), $F(1, 500) = 45.88$, $p < .001$, $\eta_p^2 = .084$. There was no main effect of ascription of drawing to revision label, $F(1, 500) = 0.98$, $p > .25$, $\eta_p = .002$, nor was there an interaction, $F(1, 500) = 1.55$, $p = .214$, $\eta_p^2 = .003$. Moreover, the revision bias was indeed moderated by the analytic-holistic reasoning scale. For each participant, we tested for moderation using the SPSS MACRO MEMORE (Montoya, 2017). As predicted, analytic-holistic scores emerged as a significant moderator ($b = .29$, $SE = .11$, $p = .008$); specifically, the revision bias was stronger among those who scored at least 4.07 on this scale relative to those who scored lower, using Johnson Neyman

significance testing. We obtained the same result when using a mixed effects model (see appendix H in Supplementary Procedure).

These findings support our hypothesis that people who are naturally more intuitive in their thinking style may be more susceptible to the revision bias. Finally, Study 5B tests for further evidence of the intuitive processing style underlying the revision bias using a method complementary: experimentally inducing thinking style via time pressure. Previous research demonstrates that time pressure causes people to adopt less effortful thinking strategies in favor of heuristics (Payne, Bettman, & Johnson, 1988; Simon, 1981). Thus, when people feel like they are under time pressure they may be more likely to exhibit the revision bias.

Study 5B: Methods

Participants (N = 406; 55.2% male; M_{age} = 34.95 years, SD = 10.07) from Amazon's MTurk completed the study, which was the first of two unrelated studies administered together for a fixed payment. Participants rated "original" and "revised versions" of a resume, via the same task and stimuli as Study 3A. As in previous studies, the ascription of font to revision label was randomized between-subjects. However, we further manipulated the absence or presence of time pressure.

Half of participants were informed that they would have unlimited time for the task. They read:

> *"You will have as much time as you want to complete this task; this means that you can spend as much time as you want on the next page reviewing and rating both drafts of the resume."*

The other half of participants faced time pressure. These participants had only 45 seconds to complete the task, and were informed of the time constraint up front. They read:

> *"You will have 45 seconds to complete this task; this means that you can spend 45 seconds maximum on the next page reviewing and rating both drafts of the resume. Please take note of the countdown clock. The screen will auto-advance after the 45 seconds are up, so you need to make sure you've answered the questions on the page by the time the screen auto-advances."*

We set the time constraint by taking the mean response time from a pilot study in which participants rated an original and a revised version of a person's resume (N = 49; 53.1% male; M_{age} = 33.51 years, SD = 9.94) with no time constraint. The mean response time was 44.9 seconds and 32.7% of participants required more than 45 seconds to complete the task. Thus, we set a time constraint of 45 seconds for the time pressure condition in the main study—a constraint that we thought would be short enough to induce a sense of time pressure but long enough such that most participants would be able to complete both outcome measures. The latter is important because it kept missing data to a minimum. 94.8% of participants completed both outcome measures, resulting in a final sample of 385 participants; perhaps not surprisingly, participants in the time pressure condition were less likely to have completed both ratings relative to those in the control condition (time pressure: 91.0% completed both; control: 98.6% completed both; $\chi^2(1) = 11.94$, $p < .001$, Cramer's V = .171). However, this proportion is very low in absolute terms.

Study 5B: Results

A 2 × 2 × 2 mixed ANOVA entering revision label, time pressure, and the ascription factor revealed the critical main effect of revision label, $F(1, 381) = 57.42$, $p < .001$, $\eta_p^2 = .131$: resumes merely labeled "revised" (M = 3.82, SD = 1.26) were more appealing than those merely labeled "original" (M = 4.21, SD = 1.32). Yet again, the revision bias emerged. Critically, we

also found the hypothesize interaction between revision label and time pressure, $F(1, 381) = 5.71$, $p = .017$, $\eta_p^2 = .015$ (see Figure 3). Specifically, the difference between the appeal of the revised versus original versions was greater under time pressure ($M_{diff} = 0.52$, $SD = 1.00$) as compared to this difference with no time pressure ($M_{diff} = 0.27$, $SD = 1.06$), $t(383) = -2.38$, $p = .018$, $d = 0.24$). Incidentally, there was also a between-subjects main effect of time pressure, $F(1, 381) = 8.98$, $p = .003$, $\eta_p^2 = .023$, revealing that ratings in the time pressure condition were overall higher than ratings in the no time pressure condition. There was no effect of the ascription factor nor did it interact with the other factors (all ps > .10).

Study 5B replicates the revision bias and also reveals when people might be more or less susceptible to it. Consistent with our proposed framework, the revision bias appears to operate as a judgment heuristic, with participants preferring revised resumes over original resumes to a greater degree when put under time pressure to evaluate those resumes (i.e., having less time to examine details and thus being more influenced by the "quick" information of the revision label).

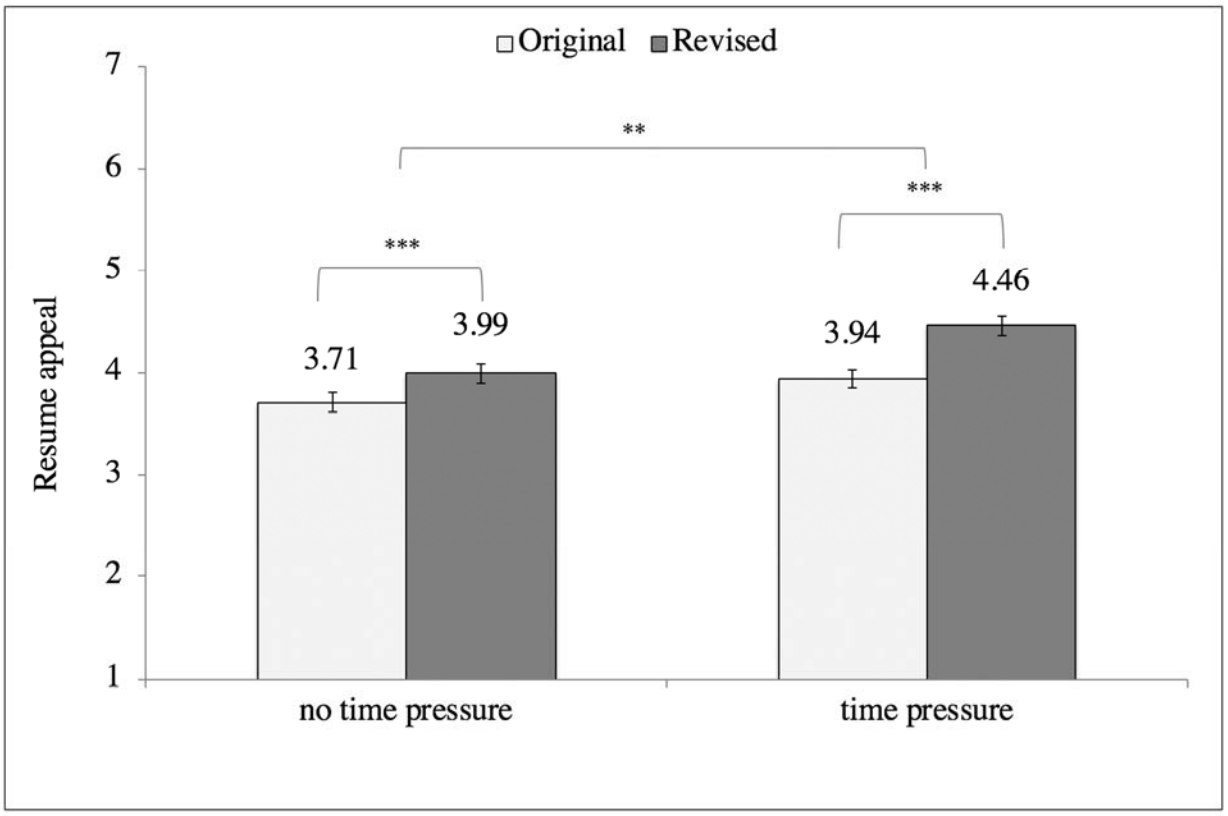

*Figure 3***.** Study 5B: mean ratings for original and revised versions across conditions. Error bars indicate ±1 SEM. Significance levels compare the two versions in each condition (*** $p < .001$; ** $p < .01$).

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 work to revise their own creations and contributions. In principle, a world of constant revision should lead to better outcomes for both creator and consumer. In practice, the current research reveals potential problems that may arise when change is not accurately advertised. Nine studies document the revision bias: people's tendency to prefer things that have been revised, regardless of whether the revised versions are objectively better than their predecessors.

Studies 1-3D showed the basic effect. In Study 1, 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. In Study 2, the same candy tasted better merely when participants thought its recipe had been revised. In Studies 3A-3D, participants preferred "revised" stimuli even when the alleged revisions were obviously trivial, incidental, identical to the original, and even inferior to the original. Next, Studies 4-5B showed how the revision bias operates as a judgment heuristic. In Study 4, the same game was more enjoyable to play when participants merely thought it was a recently revised version, driven by the mere label leading participants to be less sensitive to potential bugs while playing. In Studies 5A-5B, participants were more susceptible to the revision bias when they were engaged in a holistic processing style, whether stemming from individual differences or having been experimentally induced. Finally, our participants were diverse, drawn from student and community subject pools and national online panels, suggesting that our findings are robust and that most individuals display a preference for things that have been revised absent objective improvement.

The revision bias raises clear practical implications. On the one hand, things that are objectively unchanged (or even made worse) in the revision process may nonetheless be adopted, so long as observers *believe* they possess a "revised" version. This may happen innocently (e.g., co-authors may be prone to accepting the revisions of whoever leaves off, even if they make no objective advance), but also intentionally (e.g., companies that release annual updates for sake of releasing annual updates). As the revision bias emerged even after giving participants complete first-hand knowledge of the entity in question (e.g., actually eating a candy, actually playing a game), these possibilities raise bigger problems to the extent that people cannot easily trust their bottom-up experience to draw more informed conclusions (especially in environments where

content creators are motivated to make revisions for reasons beyond the need for revisions). On the other hand, things that are objectively improved in the revision process may nonetheless go unappreciated, if the fact that they have been revised is not made painfully clear. Indeed, the public version of a person's resume or a candy's recipe almost certainly underwent extensive editing and tweaking behind the scenes—but our studies suggest observers may not intuitively bring this to mind when encountering the final product. Job applicants may be wise to explicitly type "CV: Revised" into the header of their resume; waiters may be wise to explicitly state that the menu was "revised by the chef"; and so on. In sum, the revision bias raises dual implications for revisions wielding influence when perhaps they should not, and failing to wield influence when perhaps they should, merely due to the label that evaluators encounter beforehand.

The revision bias also makes conceptual contributions. First, our studies highlight the need for a more nuanced understanding of how heuristics may more specifically alter real-time processing. It would be less compelling if we had observed the revision bias by name only, akin to getting excited by an advertisement that promises an improved future product that has not yet been released. The revision bias is akin to seeing the advertisement but then also getting to use that product; the product is not actually improved; yet coming away believing that it is. Second, our studies extend more general psychological research on change perception, adding further evidence that people may prefer change for mere sake of change (e.g., O'Brien & Kardas, 2016)—even after they get to experience that (lack of) change first hand. Third, the revision bias joins an emerging collection of related but distinct biases that share two common 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 & Banaji, 2012; Murdock & Bennet, 1962)—a sequencing effect—research has shown that "phantom firsts"—merely framing

something as "first"—increases its appeal (LeBoeuf, Williams, & Brenner, 2014). Interestingly, these biases suggest that in the absence of explicit revision, stimuli that are framed to have occurred earlier in a sequence are preferred (Smith, Newman, & 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, & Crandall, 2010). While our results cannot be accounted for by the longevity bias—for example, in Study 3C, the third logo had existed for the same amount of time in both conditions, but framing it as revised increased preference for it—clearly more research is needed to understand when different temporal sequences are preferred. We would suggest that one critical difference between our paradigm and these other research programs is that stimuli that people believe have remain unchanged may be more prone to a bias for the original, whereas stimuli that people believe have been explicitly changed may be more prone to the revision bias.

Finally, the revision bias raises various other directions for research. First, expanding on the ideas above, future research should examine the revision bias in the context of apparent exceptions, such as when people desire older, perhaps "pre-revised," experiences (e.g., nostalgia: Wildschut, Sedikides, Arndt, & Routledge, 2006), and when people prefer original renditions of collectibles over later renditions (e.g., artwork: Newman & Bloom, 2012). We suspect that the source of the revision may prove critical; people may be more likely to show the revision bias when the same content creator releases a revised version to their own original version (as in our studies), as opposed to when a third party introduces new changes and updates from their perspective. Second, future research should assess other incentive-compatible designs similar to Study 3D. More studies should test for the revision bias when mistakenly preferring revisions is costly (although it is worth noting that, if our effect was entirely driven by a lack of incentive for

accuracy, we should have observed random preferences rather than systematic preferences). Third, future research should further examine underlying mechanisms, not least to further establish debiasing strategies that people can use to more accurately evaluate the value of "revised" products and experiences (e.g., being aware to avoid time pressure, as revealed by Study 5B). Fourth, future research could fruitfully explore other intuitions about revisions, such as whether people anticipate these effects (e.g., a team that changes its roster from season to season assuming that any change makes for positive change, without accounting for objective improvement in quality). Likewise, as perceived revision is 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 the revision bias) 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, & Hearst, 1986, heightening the revision bias).

Until these possibilities are tested, the current research warrants a closer look at the ever-revised offerings that pervade everyday life. The revision bias may often lead people astray, but knowledge of the revision bias may also spur people to better advertise their own revisions when warranted. Any author who has experienced the pain of making endless revisions to a manuscript under its fourth round of review should take heart: while one hopes that authors have genuinely improved their manuscripts, we suspect that their improvements will seem genuine regardless.

References

Agostinelli, G., Sherman, S. J., Fazio, R. H., & Hearst, E. S. (1986). Detecting and identifying change: Additions versus deletions. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 445-454. https://doi.org/10.1037/0096-1523.12.4.445

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics, 118*, 73-106. https://doi.org/10.1162/00335530360535153

Barasz, K., John, L. K., Keenan, E. A., & Norton, M. I. (2017). Pseudo-set framing. *Journal of Experimental Psychology: General, 146*, 1460-1477. https://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1037/xge0000337

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics, 8*, 279-286. https://doi.org/10.3758/bf03212593

Bless, H., & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. *Advances in Experimental Social Psychology, 42*, 319-373. https://doi.org/10.1016/S0065-2601(10)42006-7

Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychological Bulletin, 106*, 265-289. https://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1037/0033-2909.106.2.265

Buell, R. W., & Norton, M. I. (2011). The labor illusion: How operational transparency increases perceived value. *Management Science, 57*, 1564-1579. https://doi.org/10.1287/mnsc.1110.1376

Carney, D. R., & Banaji, M. R. (2012). First is best. *PloS ONE, 7*, e35088. https://doi.org/10.1371/journal.pone.0035088

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*, 752-766. https://doi.org/10.1037/0022-3514.39.5.752

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66*, 460-473. https://doi.org/10.1037/0022-3514.66.3.460

Chinander, K. R., & Schweitzer, M. E. (2003). The input bias: The misuse of input information in judgments of outcomes. *Organizational Behavior and Human Decision Processes, 91*, 243-253. https://doi.org/http://dx.doi.org/10.1016/S0749-5978(03)00025-6

Choi, I., Koo, M., & Choi, J. A. (2007). Individual differences in analytic versus holistic thinking. *Personality and Social Psychology Bulletin, 33*, 691-705. https://doi.org/10.1177/0146167206298568

Eidelman, S., Pattershall, J., & Crandall, C. S. (2010). Longer is better. *Journal of Experimental Social Psychology, 46*, 993-998. https://doi.org/10.1016/j.jesp.2010.07.008

Furnham, A. (1984). Many sides of the coin: The psychology of money usage. *Personality and Individual Differences, 5*, 501-509. https://doi.org/10.1016/0191-8869(84)90025-4

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics, 40*, 35-42. https://dx.doi.org/10.1016/j.socec.2010.10.008

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. *Cognitive Psychology, 38*, 79-109. https://doi.org/10.1006/cogp.1998.0708

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles.* (pp. 133-168). New York, NY, US: Guilford Press.

Hsee, C. K., & Abelson, R. P. (1991). Velocity relation: Satisfaction as a function of the first derivative of outcome over time. *Journal of Personality and Social Psychology, 60*, 341-347. https://doi.org/10.1037/0022-3514.60.3.341

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science, 5*, 343-355. https://doi.org/10.1177/1745691610374586

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus, and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251. https://doi.org/10.1037/h0034747

Klaaren, K. J., Hodges, S. D., & Wilson, T. D. (1994). The role of affective expectations in subjective experience and decision-making. *Social Cognition, 12*, 77-101. https://doi.org/10.1521/soco.1994.12.2.77

Klein, N., & O'Brien, E. (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*, 210-229. https://doi.org/10.1037/pspa0000088

Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). The effort heuristic. *Journal of Experimental Social Psychology, 40*, 91-98. https://doi.org/10.1016/S0022-1031(03)00065-9

LeBoeuf, R. A., Williams, E. F., & Brenner, L. A. (2014). Forceful phantom firsts: Framing experiences as firsts amplifies their influence on judgment. *Journal of Marketing Research, 51*, 420-432. https://doi.org/10.1509/jmr.11.0298

Lee, L., Frederick, S., & Ariely, D. (2006). Try it, you'll like it: The influence of expectation, consumption, and revelation on preferences for beer. *Psychological Science, 17*, 1054-1058. https://doi.org/10.1111/j.1467-9280.2006.01829.x

Loewenstein, G. F., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review, 100*, 91-108. https://doi.org/10.1037/0033-295X.100.1.91

Mochon, D., & Frederick, S. (2013). Anchoring in sequential judgments. *Organizational Behavior and Human Decision Processes, 122*, 69-79. https://doi.org/10.1016/j.obhdp.2013.04.002

Montoya, A. K. (2017). *Simple slopes and Johnson-Neyman probing methods extended to two-condition within-subject designs.* Paper presented at the Annual Convention of the Association for Psychological Science, Boston, MA.

Morales, A. C. (2005). Giving firms an "E" for effort: Consumer responses to high-effort firms. *Journal of Consumer Research, 31*, 806-812. https://doi.org/10.1086/426615

Murdock, J., & Bennet, B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*, 482-488. https://dx.doi.org/10.1037/h0045106

Newman, G. E., & Bloom, P. (2012). Art and authenticity: The importance of originals in judgments of value. *Journal of Experimental Psychology: General, 141*, 558-569. https://doi.org/10.1037/a0026035

O'Brien, E., & Kardas, M. (2016). The implicit meaning of (my) change. *Journal of Personality and Social Psychology, 111*, 882-894. https://doi.org/10.1037/pspi0000073

O'Brien, E., & Ellsworth, P. C. (2012). More than skin deep: Visceral states are not projected onto dissimilar others. *Psychological Science, 23*, 391-396. https://doi.org/10.1177/0956797611432179

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 534-552. https://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1037/0278-7393.14.3.534

Ryff, C. D. (1991). Possible selves in adulthood and old age: A tale of shifting horizons. *Psychology and Aging, 6*, 286-295. https://doi.org/10.1037/0882-7974.6.2.286

Sheldon, K. M., Boehm, J. K., & Lyubomirsky, S. (2012). Variety is the spice of happiness: The hedonic adaptation prevention (HAP) model. *Oxford handbook of happiness*, 901-914.

Sher, S., & McKenzie, C. R. (2011). Levels of information: A framing hierarchy. In G. Keren (Ed.), *Perspectives on framing* (pp. 35-64). New York, NY: Psychology Press.

Simon, H. A. (1957). Models of man, social and rational; mathematical essays on rational human behavior in society setting. *The Journal of Philosophy, 59*, 177-182.

Simon, H. A. (1981). *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Smith, R. K., Newman, G. E., & Dhar, R. (2015). Closer to the creator: Temporal contagion explains the preference for earlier serial numbers. *Journal of Consumer Research, 42*, 653-668. https://doi.org/10.1093/jcr/ucv054

TIME. (2017). Jimmy Kimmel pranks clueless people into thinking the iPhone 4 is the iPhone X. In B. Perrigo (Ed.), *TIME*.

Tormala, Z. L., Jia, J. S., & Norton, M. I. (2012). The preference for potential. *Journal of Personality and Social Psychology, 103*, 567-583. https://doi.org/10.1037/a0029227

Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In W. Derek & C. Vlek (Eds.), *Utility, Probability, and Human Decision Making* (Vol. 11, pp. 141-162). Theory and Decision Library: Springer, Dordrecht.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453-458. https://doi.org/10.1126/science.7455683

Wildschut, T., Sedikides, C., Arndt, J., & Routledge, C. (2006). Nostalgia: Content, triggers, functions. *Journal of Personality and Social Psychology, 91*, 975-993. https://doi.org/10.1037/0022-3514.91.5.975

Wilson, A. E., & Ross, M. (2001). From chump to champ: People's appraisals of their earlier and present selves. *Journal of Personality and Social Psychology, 80*, 572-584. https://doi.org/10.1037/0022-3514.80.4.572

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1-27. https://doi.org/10.1037/h0025848