How Does Rating Specific Features of an Experience Alter Consumers' Overall Evaluations of That Experience?

Katie S. Mehr

Joseph P. Simmons

Katie S. Mehr (kmehr@ualberta.ca) is an Assistant Professor of Marketing, University of

Alberta, Edmonton, AB T6G 2R6. Joseph P. Simmons (jsimmo@wharton.upenn.edu) is the

Dorothy Silberberg Professor of Applied Statistics and a Professor of Operations, Information,

and Decisions, University of Pennsylvania, Philadelphia, PA, 19146. Address all correspondence

to Katie S. Mehr. This paper is based on the lead author's dissertation completed under the

supervision of the second author.

ABSTRACT

How does the way companies elicit ratings from consumers affect the ratings that they receive? In 10 pre-registered experiments, we find that consumers rate subpar experiences more positively overall when they are also asked to rate specific aspects of those experiences (e.g., a restaurant's food, service, and ambiance). Studies 1-4 established the basic effect across different scenarios and experiences. Study 5 found that the effect is limited to being asked to *rate* specific features of an experience, rather than providing open-ended comments about those features. Studies 6-9 provided evidence that the effect does not emerge because rating positive aspects of a subpar experience reminds consumers that their experiences had some good features. Rather, it emerges because consumers want to avoid incorporating negative information into both the overall and the attribute ratings. Lastly, study 10 found that asking consumers to rate attributes of a subpar experience reduces the predictive validity of their overall rating. We discuss implications of this work and reconcile it with conflicting findings in the literature.

Keywords: ratings, feedback, evaluations

Companies often ask consumers to provide feedback. For instance, after taking an Uber, you will likely receive a push notification asking you to rate your ride. Or, after dining at a restaurant booked on OpenTable, you may receive an email asking you to rate your experience. But companies vary in how they elicit this feedback. For instance, whereas Uber asks its customers to provide only one overall rating, OpenTable asks its customers to provide an overall rating and to rate specific aspects of their experience, such as the restaurant's food, service, ambiance, and value. In this research, we shed light on how these different elicitation methods affect consumers' ratings. Specifically, we investigate how being asked to rate multiple, specific aspects of an experience (e.g., a restaurant's food, service, ambiance, and value) affects consumers' overall evaluations of that experience.

Across ten pre-registered experiments, we find that asking consumers to rate specific attributes of a subpar experience (i.e., an experience with some desirable and undesirable aspects) increases their overall evaluation of that experience. For example, we find that consumers who experience bad service at a restaurant provide higher overall ratings of that experience when they are also asked to rate the restaurant on food, service, ambiance, and value. We find that this effect arises because being able to rate a specific negative aspect of an experience makes consumers less inclined to incorporate that negative aspect into their overall evaluation.

This research makes both theoretical and practical contributions. Theoretically, we contribute to a large literature investigating how consumers evaluate their experiences, shedding light on how, why, and when those evaluations may be influenced by contextual factors. Practically, our research suggests that asking consumers to rate multiple aspects of their subpar experiences prompts them to provide overall ratings that are more positive but less accurate/predictive. So whether firms should ask consumers to rate multiple aspects of their

experiences depends on whether those firms want to bolster the positivity or the accuracy of consumers' overall ratings.

CONCEPTUAL BACKGROUND

Service Quality and Online Consumer Ratings

Consumers often share feedback about a product they bought or an experience they had. A large literature on service quality has investigated how various factors affect what feedback consumers provide, and shed light on how to measure consumers' satisfaction with services (Asubonteng, McCleary and Swan 1996; Collier and Bienstock 2006; Cronin Jr and Taylor 1992; Landrum et al. 2008; Messinger et al. 2009; Santos 2003; Seth, Deshmukh and Vrat 2005; Sivadas and Baker-Prewitt 2000). For instance, this literature has developed rating scales to use to understand consumers' perceptions of service quality (Cronin Jr and Taylor 1992; Lytle, Hom and Mokwa 1998; Parasuraman, Zeithaml and Berry 1988). Researchers have studied these scales in a variety of contexts (Bahia and Nantel 2000; Bienstock, Mentzer and Bird 1997; Brown and Swartz 1989; Parasuraman, Zeithaml and Malhotra 2005; Pitt, Watson and Kavan 1995; Urden 2002) and investigated whether different measures of satisfaction (e.g., 5 vs. 10-point scales) affect consumers' reported satisfaction (Coelho and Esteves 2007). While some work in this area has explored how various dimensions affect service quality (Kang and James 2004; Lehtinen and Lehtinen 1991; Santos 2003), no research has experimentally investigated how the inclusion of attribute ratings affects consumers' overall ratings.

Consumers increasingly share their feedback by providing online ratings of products and experiences on websites like Yelp, OpenTable, and TripAdvisor. As online ratings have become ubiquitous, a large and growing body of research has focused on how online ratings affect consumers' judgments and decisions. From this literature we know that consumers often use ratings, that ratings affect sales, and that companies often encourage consumers to provide ratings (Burtch et al. 2018; Chevalier and Mayzlin 2006; Chintagunta, Gopinath and Venkataraman 2010; Donati 2022; Goes, Lin and Au Yeung 2014; He and Bond 2015; Kwark, Chen and Raghunathan 2014; Simonson 2016; Zhu and Zhang 2010).

The literature also sheds light on how consumers process product ratings and reviews. For example, research suggests that consumers' choices and evaluations are influenced by the mean, variance, and number of consumer ratings (Chevalier and Mayzlin 2006; Chu, Roh and Park 2015; Clemons, Gao and Hitt 2006; He and Bond 2015; Khare, Labrecque and Asare 2011; West and Broniarczyk 1998). More specifically, research suggests that consumers are very attentive to average ratings while being insufficiently attentive to other aspects of those ratings, such as the number of ratings on which that average is based (de Langhe, Fernbach and Lichtenstein 2016). Research also suggests that consumers make categorical distinctions between positive and negative reviews, while being insufficiently sensitive to differences between reviews of the same valence (Fisher, Newman and Dhar 2018). Finally, we also have some sense of which reviews consumers find most helpful, although exactly which variables predict helpfulness varies across product types (Li et al. 2013; Mudambi and Schuff 2010; Pan and Zhang 2011).

Consumers' heavy reliance on online ratings means that it is important to understand not only how they interpret those ratings, but also how they generate them. Extant research on this question has uncovered several factors that may influence consumers' ratings and reviews,

including exposure to reviews provided by experts (Jacobsen 2015) or other consumers (Moe and Schweidel 2012; Sridhar and Srinivasan 2012), precisely how the rating scales are labeled (Jiang and Guo 2015; Tsekouras 2017), whether the reviews are oral or written (Berger and Iyengar 2013), and whether they are done on a smartphone or computer (Melumad, Inman and Pham 2019).

Nevertheless, despite these contributions, there is still much we do not know about how consumers generate ratings of their experiences. For example, we do not have a thorough understanding of how the way in which consumers' evaluations are elicited affects the ratings that they provide. This is important because companies vary widely in how they elicit ratings. For example, while Airbnb and OpenTable ask consumers to rate various aspects of their stay or dining experience, sites like Uber only ask for an overall rating. Some other sites, like TripAdvisor or Etsy, prompt consumers to write a written review and provide an overall rating.

Attitudes and Attitude Measurement

Consumers can provide online ratings to express how positively or negatively they feel about a given product or experience, or in other words, their attitudes. Previous literature has described an attitude as "a summary evaluation of a psychological object captured in such attribute dimensions as good-bad, harmful-beneficial, pleasant-unpleasant, and likeable-dislikeable" (Ajzen 2001, 29). Consumers often form attitudes by learning about a product or experience and using that information to form an evaluation (Argyriou and Melewar 2011; Fazio, Lenn and Effrein 1984; Lutz 1991; Van Overwalle and Siebler 2005). For instance, if there is a new restaurant in town, consumers learn about the restaurant through advertisements they see, word of mouth from family and friends, and their direct experiences with the restaurant. This information helps consumers

form beliefs about different aspects of the experience, and these beliefs can be aggregated to form an attitude (Ajzen 2001).

Marketing researchers, business owners, and other entities often want to assess consumers' attitudes. Although there are many ways to do this, probably the most straightforward way is to ask consumers to directly report them. But there is an important complication, which is that contextual factors can affect consumers' responses. Thus, the responses researchers receive from consumers when asking for their attitudes will be a function of not only the attitudes consumers have, but also a number of contextual factors. For example, previous work suggests that consumers' reported attitudes can be influenced by the question being asked (Bickart, Phillips and Blair 2006; Gal and Rucker 2011), the response scale (Krosnick and Alwin 1987; Nowlis, Kahn and Dhar 2002; Weijters and Baumgartner 2012; Weijters, Cabooter and Schillewaert 2010), and the ordering of the questions (Bickart 1993; DeMoranville and Bienstock 2003; Peterson and Wilson 1992; Schwarz 1999; Schwarz and Bless 1992; Schwarz and Strack 1991; Schwarz, Strack and Mai 1991; Simmons, Bickart and Lynch Jr 1993; Tourangeau and Rasinski 1988). Research suggests that contextual factors are more likely to influence consumers' responses when consumers' attitudes are more weakly or less certainly held than when they are more strongly or more certainly held (Tormala and Rucker 2007). For instance, when consumers are more certain of their attitudes, the attitudes are more durable and thus less affected by context effects (Tormala and Rucker 2018).

How Might the Presence of Attribute Ratings Change Consumers' Overall Ratings?

In this paper, we consider how one specific contextual factor affects consumers' reported attitudes: the presence of attribute ratings. Specifically, we investigate how being asked to rate multiple, specific attributes of an experience (e.g., a restaurant's food, service, ambiance, and value) affects consumers' overall evaluations of that experience.

In ten experiments, we find that asking consumers to rate specific attributes of a subpar experience (i.e., an experience or product with some bad aspects) increases their overall evaluation of that experience. For example, consumers who experience bad service at a restaurant provide higher overall ratings of that experience when they are also asked to rate the restaurant on food, service, ambiance, etc. This effect does not emerge for uniformly good experiences. We consider and test competing explanations for this effect: "positive reminders" and "avoiding negative redundancy."

Positive Reminders. First, it is possible that asking consumers to rate multiple attributes of a subpar experience reminds them that many aspects of their experience were actually quite satisfactory (Earthy, MacFie and Hedderley 1997). An important literature on negativity bias suggests that consumers often attend more and give greater weight to negative stimuli than to positive stimuli (Baumeister et al. 2001; Fiske 1981; Rozin and Royzman 2001). The presence of negativity bias would suggest that when a consumer goes to a restaurant with bad food and good service, the bad food is likely to be quite salient precisely because it is negative. Thus, if consumers are asked only to rate the overall experience, consumers might focus on the bad aspect (e.g., the food), and give it considerable weight in their resulting evaluation. As a consequence, their overall rating may be quite poor.

But what if consumers are asked to provide not only an overall evaluation of their experience, but also to rate specific aspects of that experience, some of which are good (the service) and some of which are bad (the food)? It is possible that asking them to rate multiple aspects might draw their focus away from the bad aspects and turn some of their attention to the good aspects. Indeed, rating multiple aspects of an experience might remind consumers that the experience was not all bad, and thus lead them to incorporate the good aspect(s) of their experience more heavily in their overall rating, resulting in a more positive overall rating. Notably, if a consumer has a purely good experience, then there are not any negative aspects to focus on, and so a reminder of the positive aspects of the experience should not be consequential (as those aspects are already salient). Thus, this "positive reminders" account predicts that the presence of attribute ratings will affect the overall evaluation consumers provide for subpar experiences, but not for good experiences.

A recent paper provides some evidence for this account, or at least for one that is very similar to it (Schneider et al. 2020). These authors propose that consumers' overall ratings reflect an average of the specific attribute ratings that they provide, so that asking consumers to rate more positive attributes produces more positive overall evaluations than asking them to rate more negative attributes.

Avoiding Negative Redundancy. When providing feedback, many consumers will feel compelled to incorporate the negative aspects of their experiences. For example, many restaurant customers who experience bad food and good service will feel compelled to give lower overall evaluations than those who experience good food and good service. Put simply, most consumers want to be truthful. At the same time, many consumers may prefer to provide ratings that err on

the side of generosity, so to avoid giving overly harsh feedback. Indeed, most customer reviews are positive (Chevalier and Mayzlin 2006), and there is evidence that this arises in part because consumers are more inclined to withhold negative feedback than positive feedback (Berg et al. 2020). This desire may reflect people's preference to be nice or non-negative, to think of themselves as nice people (Berger 2014; Chung and Darke 2006; Hennig-Thurau et al. 2004; Sundaram, Mitra and Webster 1998), and/or to provide ratings that have less impact, as negative reviews can be especially consequential (and therefore damaging; Chevalier and Mayzlin 2006). In sum, consumers may prefer to be truthful, while erring on the side of being charitable.

If restaurant consumers experience bad food and good service, but are only asked to provide a single, overall rating, then they cannot be truthful and charitable at the same time. They can either truthfully provide a rating that fully communicates the subpar nature of their experience, or they can artificially inflate their rating. But if consumers are asked to provide an overall rating *and* to rate the specific attributes of their experience (e.g., the food and the service), then they have an opportunity to be both truthful and charitable. They can truthfully rate the food to be subpar, while also giving that aspect less weight in their overall rating, thereby inflating it. Indeed, many consumers may feel bad about providing two negative ratings – for both the food and the overall experience – when providing one negative rating sufficiently communicates the negative aspect of their experience (Grice 1975). Why be negative twice when you can be (more specifically) negative once?

This account has some support in research on complaining behavior (Bennett 1997; Blodgett, Wakefield and Barnes 1995; Halstead and Page Jr 1992; Hunt 1991; Nyer 1999, 2000), which suggests that when consumers provide higher satisfaction ratings when they have the opportunity to complain about a negative service interaction. However, this research has found

that this effect emerges only when there is a time delay (usually around 2 weeks) between the complaint and the satisfaction rating (Nyer 1999, 2000). Because there is no time delay in the contexts we investigate, this research would not predict the effects that we observe.

Both the "positive reminders" and the "avoiding negative redundancy" accounts predict that consumers will provide more positive ratings of subpar experiences when they are also given the opportunity to rate specific aspects of those experiences. But they make different predictions about why and when this effect will emerge. Whereas the "positive reminders" account suggests that rating *positive* aspects of a subpar experience will increase consumers' overall ratings, since the act of rating positive aspects will serve as a reminder that the experience was not so bad, the "avoiding negative redundancy" account suggests that rating *negative* aspects will increase consumers' overall ratings, since consumers may feel compelled to give more positive overall ratings when they can specifically rate the negative aspects of that experience. As presented below, our evidence favors the "avoiding negative redundancy" account.

OVERVIEW OF STUDIES

In this paper, we report 10 pre-registered experiments investigating whether asking consumers to directly rate the specific aspects of their experiences alters how they rate those experiences overall (see table 1). In studies 1-4, we find, across many different scenarios and experiences, that rating specific aspects of a subpar experience increases overall ratings of those experiences. We find no effect for "good" experiences. We then investigate whether this effect also emerges when consumers have the opportunity to say specific things about their experiences in an open-ended text box (study 5). In studies 6-9, we test two competing categories of mechanisms for

the effect. Lastly, study 10 examines whether asking consumers to rate specific attributes of subpar experiences increases or decreases the predictive value of their overall ratings. All of our code, data, materials, and pre-registrations can be found on ResearchBox: https://researchbox.org/196.

TABLE 1

OVERVIEW OF STUDIES

Study	Rating Conditions	Study Domain	Experience Type	Rating Scale	Key Finding
1	Overall only, overall + attributes	Restaurants	Hypothetical	5 stars	Rating specific aspects of subpar experiences (but not good experiences) increases consumers' overall ratings of that experience.
2	Overall only, overall + attributes	Airbnb Rentals	Hypothetical	5 stars	
3	Overall only, overall + attributes	Restaurants and ride shares	Recalled	5 stars	Rating specific aspects of (recalled or real) subpar experiences increases consumers' overall ratings of that experience.
4	Overall only, overall + individual paintings, overall + painting attributes	Painting gallery	Real	5 stars	
5	Overall only, overall + attributes, overall + text box	Airbnb Rentals	Hypothetical	5 stars	Rating specific aspects of subpar experiences increases consumers' overall ratings of that experience, but providing feedback in a text box does not.
6	Overall only, overall + attributes, overall + negative attributes	Airbnb Rentals	Hypothetical	5 stars	Rating specific negative aspects of subpar experiences increases consumers' overall ratings of that experience.
7	Overall only, overall + attributes, overall + positive attributes	Airbnb Rentals	Hypothetical	5 stars	Rating specific positive aspects of subpar experiences does not increase consumers' overall ratings of that experience.
8	Overall only, overall + attributes, overall + positive attributes, overall + negative attributes	Restaurants	Hypothetical	5 stars	Rating specific negative aspects of subpar experiences increases consumers' overall ratings of that experience, but ratings specific positive aspects does not.
9	Overall only, overall + attributes, overall + positive attributes, overall + negative attributes	Commercial flights, dentist office visits, bookcases, sneakers	Hypothetical	10 stars	
10	Overall only, overall + attributes	Restaurants	Hypothetical	10 stars	Rating specific aspects of subpar experiences decreases the predictive validity of consumers' overall ratings of that experience.

NOTE. – This table provides an overview of the studies in this manuscript.

STUDIES 1 AND 2

In this paper, we hypothesize that consumers will rate experiences more positively overall when they are given the opportunity to directly evaluate the negative aspects of those experiences.

In studies 1 and 2, we tested this hypothesis by examining whether asking consumers to rate several specific attributes of an experience increases their overall evaluations of that experience, and whether this effect emerges only for experiences that actually had negative aspects.

In each study, participants read two scenarios describing either a restaurant experience (study 1) or a stay at an Airbnb (study 2). One scenario asked them to imagine a uniformly good experience, in which everything went at least as well as expected, and one asked them to imagine a subpar experience, in which at least one thing was worse than expected. For each scenario, participants in the "overall + attributes" condition rated their overall experience as well as several attributes of that experience, whereas those in the "overall only" condition rated their overall experience and nothing else. We expected overall ratings to be higher in the "overall + attributes" condition when the experience was subpar. However, we did not expect to observe this for "good" experiences, for which there was no tension between being truthful and charitable.

Method

Participants. For both studies, we recruited online participants from Amazon's Mechanical Turk (MTurk). Participants in study 1 were paid $0.40, while those in study 2 were paid $0.30. We pre-registered to collect data from 1,000 participants and to include only the first response in the event of a duplicate IP address or MTurk ID (All pre-registrations can be found in our ResearchBox: https://researchbox.org/196). For study 1, we removed the responses of three duplicate participants, as well as seven observations from four participants who failed to complete the dependent variable; our final sample consisted of 1,001 participants and 2,001 observations (mean age = 39.23, 52.75% female). For study 2, we removed the responses of four duplicate

participants, as well as 28 observations from 15 participants who failed to complete the dependent variable; our final sample consisted of 1,003 participants and 2,004 observations (mean age = 38.14, 51.74% female).

Procedure. Participants began the survey by reading a set of instructions.[1] On each of the next two pages, participants read a scenario in which they imagined having a particular experience, and then rated the quality of that experience. In study 1, both scenarios were about a restaurant experience. In study 2, both scenarios were about an Airbnb stay.

FIGURE 1

RATING QUESTIONS FOR EACH CONDITION IN STUDY 2

We manipulated scenario quality within-subjects, as all participants read one scenario that described a good experience (i.e., pretested to be around 4 out of 5 stars), and one scenario that

---

[1] A typo in the instructions of studies 1 and 2 told participants in both conditions that they would be indicating whether they would provide a rating. In reality, they were asked what rating they would provide.

described a subpar experience (i.e., pretested to be around 2 out of 5 stars). In study 1 participants saw two scenarios randomly drawn from a pool of four possible scenarios (i.e., 2 good scenarios and 2 subpar scenarios), while in study 2 participants saw two scenarios randomly drawn from a pool of six possible scenarios (i.e., 3 good scenarios and 3 subpar scenarios). We randomized the order of the good and subpar scenarios across participants.

Between subjects, participants were randomly assigned to either the *overall only* condition or the *overall + attributes* condition. In the *overall only* condition, participants simply rated the quality of their overall experience after reading each scenario. In the *overall + attributes* condition, participants rated the quality of their overall experience followed by several attributes for each of the two scenarios. In study 1, these attributes were food, service, ambiance, and value. In study 2, these attributes were accuracy, check-in, cleanliness, communication, location, and value. Ratings were made on a one to five-star scale and could only be whole numbers.[2] Participants' rating of their overall experience was our main dependent variable. Figure 1 shows what our manipulation and measures looked like in study 2. After rating each of the two scenarios, participants reported their age and gender.

Results and Discussion

Figure 2 displays the mean overall rating by condition. Each participant contributed two observations to the dataset, one for the subpar scenario and one for the good scenario. We regressed the overall rating on a contrast-coded indicator for the *overall + attributes* condition (-0.5 = *overall*

---

[2] We did not realize this in advance, but the ratings interface on Qualtrics (our survey software) allows participants to type in a rating of 0 after clicking on one of the five-star ratings. Three participants entered overall ratings of zero in study 1. Because we did not pre-register to exclude ratings of zero stars, these ratings were retained in our analyses. Excluding them does not affect the significance of any of our results.

only; 0.5 = *overall + attributes*), a contrast-coded indicator for whether the scenario was subpar (-0.5 = *good*; 0.5 = *subpar*), the interaction of these two indicators, and fixed effects for scenario. To account for the nonindependence of observations by the same participant, we clustered standard errors by participant. We predicted that participants in the *overall + attributes* condition would rate the overall experience higher than those in the *overall only* condition, but only when the experience was subpar. In other words, we expected a significant interaction between the *overall + attributes* condition and whether the scenario was subpar. This interaction materialized in both studies (study 1: b = .24, SE = .07, $t(1000)$= 3.35, p < .001; study 2: b = .27, SE = .08, $t(1002)$= 3.44, p < .001).

FIGURE 2

STUDIES 1 AND 2 RESULTS
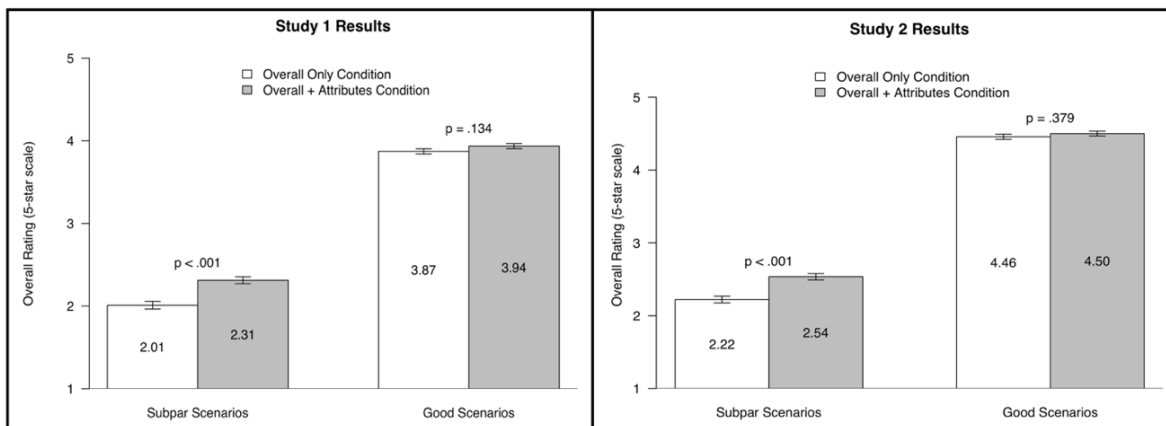
NOTE. - Error bars depict +/- 1 standard error.

To unpack this interaction, we examined the effect of the *overall + attributes* vs. *overall only condition* separately for the subpar and good scenarios. As predicted, whereas participants in the *overall + attributes* condition rated the overall experience more positively than those in the

overall only condition when the experience was subpar (study 1: $b = .30$, SE $= .06$, $t(997)= 4.81$, $p < .001$; study 2: $b = .31$, SE $= .06$, $t(997)= 5.06$, $p < .001$), there was no significant difference between the *overall + attributes* condition and the *overall only* condition when the experience was good (study 1: $b = .06$, SE $= .04$, $t(998) = 1.50$, $p = .134$; study 2: $b = .04$, SE $= .05$, $t(999)= .88$, $p = .379$).

The results of studies 1 and 2 provide initial evidence for our main hypothesis. Consumers rate subpar experiences more positively overall when they are asked to rate multiple aspects of those experiences. This does not seem to happen when the experiences are good.

STUDY 3

Studies 1 and 2 provide some initial evidence that consumers give higher overall ratings to subpar experiences when they are asked to rate the overall experience in addition to several attributes of the experience (compared to rating the overall experience alone). In study 3, we sought to go beyond these hypothetical scenario studies to investigate whether the effect would generalize to circumstances in which participants were asked to recall and rate a subpar experience that they recently had.

Method

Participants. We recruited online participants from Prolific Academic and paid them $0.45. We pre-registered to collect data from 2,000 participants and to exclude all duplicate responses (28 participants), responses from any Prolific IDs that were not submitted for payment (34

participants), and any participants who indicated that they had not had one of the experiences we specified (273 participants). We also necessarily removed 133 participants who failed to complete the main dependent variable. Our final sample consisted of 1,944 participants and observations (mean age = 40.15, 52.78% female).

Procedure. On the first page of the survey after the consent form, participants read the following: "In this survey, you'll be asked several questions about an experience you had. Which of the following experiences have you had that you can remember? If you can remember more than one, pick the most recent one." Participants could select one of the following options in response to this question: a restaurant experience with bad food and good service; a restaurant experience with bad service and good food; a ride share experience with a friendly but unsafe driver; a ride share experience with a safe but unfriendly driver; I can't remember any experiences like the ones described above. If participants selected that they could not remember any experiences like the ones described, then they were redirected to the end of the survey, instructed not to try completing the survey again, and did not receive payment. On the next few pages, participants who could remember a relevant experience answered questions about the experience (e.g., what was the name of the restaurant you went to?).

As in studies 1 and 2, we manipulated rating condition between-subjects. Participants were randomly assigned to either the *overall only* condition or the *overall + attributes* condition. For the restaurant experiences, participants in the *overall + attributes* condition rated the experience on food, service, ambiance, and value, in addition to providing an overall rating. For the ride share experiences, participants in the *overall + attributes* condition rated the driving quality, driver quality, vehicle quality, and navigation/route and pick-up/drop-off locations, in addition to

providing an overall rating. In the *overall + attributes* condition, the overall rating was listed first, followed by the ratings of the attributes. For both types of experiences, participants in the *overall only* condition provided only an overall rating. As in studies 1 and 2, our primary dependent variable was the overall rating, indicated on a 5-star rating scale. After providing a rating, participants reported their age and gender.

Results and Discussion

FIGURE 3

MEAN OVERALL RATINGS BY CONDITION AND EXPERIENCE FOR STUDY 3
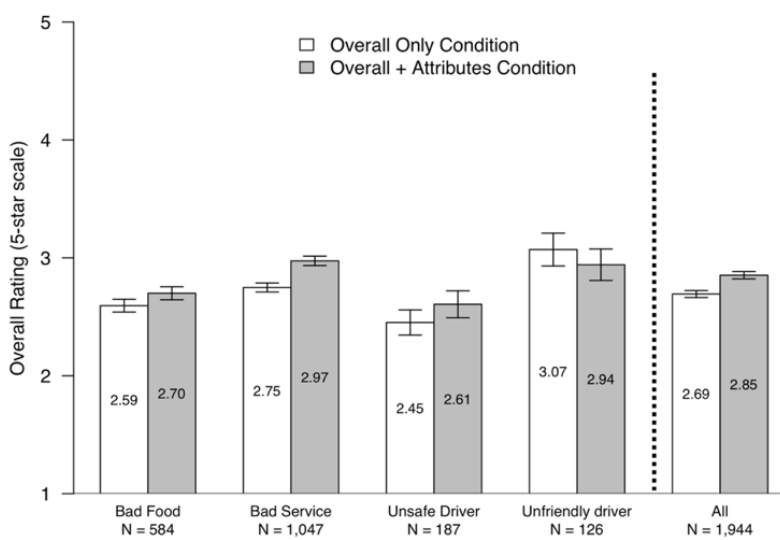
NOTE. - Error bars depict +/- 1 standard error.

To test whether participants rated the overall experience more positively when they were also asked to rate specific aspects of the experience, we regressed their overall rating on an indicator for the *overall + attributes* condition (-0.5 = *overall only*; 0.5 = *overall + attributes*) and fixed effects for experience. Participants in the *overall + attributes* condition once again gave more

positive overall ratings (M = 2.85, SD = 0.96) than those in the *overall only condition* (M = 2.69, SD = 0.93; b = .16, SE = .04, t(1939) = 3.75, p < .001), suggesting that the effect we observed in studies 1 and 2 is robust to recalled experiences.[3]

STUDY 4

Study 4 had two objectives. First, and most importantly, we wanted to examine whether our effect would emerge when participants actually underwent an experience rather than reading about one or recalling one. To achieve this, we asked participants in study 4 to view a gallery of four paintings and to rate the overall quality of the gallery. Second, we were interested in examining whether the effect is robust to different operationalizations of the *overall + attributes* condition. In the *overall + individual paintings* condition, participants provided an overall rating of the painting gallery, as well as their ratings of the individual paintings that made up that gallery. In the *overall + painting attributes* condition, participants provided an overall rating of the painting gallery, as well as their ratings of the attributes of those paintings (e.g., creativity, originality, etc.).

Method

Participants. We recruited online participants from Prolific Academic and paid them $0.50. We pre-registered to collect data from 1,500 participants and to exclude all duplicate responses (10

---

[3] As shown in figure 3, there is some apparent heterogeneity in the effects we found in study 3. Specifically, we found our directional effect among participants who indicated that the remembered an experience with bad food (d = .11, t(582) = 1.36, p = .18) , with bad service (d = .25, t(1045) = 4.08, p < .001), and with an unsafe driver (d = .15, t(185) = .99, p = .32). However, the effect directionally (but not significantly) reversed for the 10.0% of participants who indicated that they had an unfriendly driver (d = -.12, t(124) = .66, p = .51). Despite this apparent heterogeneity, an omnibus F test revealed no significant condition x experience type interaction (F(3, 1936) = 1.58, p = .192), indicating that there is no significant evidence that the effect differed by type of experience.

participants), responses from any Prolific IDs that were not submitted for payment (9 participants), and any participants who failed an attention check question that occurred before random assignment (12 participants). We also removed two participants who did not complete the dependent variable. Our final sample consisted of 1,481 participants (mean age = 34.05, 46.66% female).

Procedure. Prior to being randomly assigned to condition, participants were told, "In this survey, you'll see a gallery of 4 paintings. You'll then rate the gallery." After completing an easy attention check question,[4] participants moved on to a separate page on which they were shown a screenshot of the rating question(s) that they would be asked after viewing the gallery of paintings.

The rating question(s) differed by condition. In the *overall only condition*, participants provided only an overall rating of the painting gallery. In the *overall + individual paintings* condition, participants provided an overall rating as well as a rating of each painting. And in the *overall + painting attributes* condition, participants provided an overall rating as well as a rating of the paintings on the following attributes: Creativity, Skill & technique, Clarity of themes, and Originality. All ratings were made on a 5-star scale. In both versions of the *overall + attributes* conditions, the overall rating was listed first, followed by the ratings of the paintings/attributes.

After viewing the screenshot that displayed what their upcoming task would entail, participants viewed a subpar gallery of four paintings, comprised of three that were mediocre and

---

[4] The question asked participants, "Based on the description above, what will you do in this survey?" The response options were (1) See and rate a painting gallery, (2) Watch a video about the history of Thanksgiving, (3) Identify animals in photos, (4) Donate to charity, and (5) Play a crossword game.

one that was pretty bad,[5] and then provided their rating(s). Lastly, participants reported their age and gender.

Results and Discussion

As in studies 1-3, participants' overall evaluations of the (subpar) painting gallery were more positive when they rated specific features of the gallery than when they just provided an overall rating. This effect was highly significant when comparing the *overall only* condition ($M =$ 3.24, SD = 0.96) to the *overall + painting attributes* condition ($M = 3.43$, SD = 0.84; b = .20, SE = .06, $t(1478) = 3.60$, $p < .001$), and marginally significant when comparing the *overall only* condition to the *overall + individual paintings* condition ($M = 3.34$, SD = 0.79; b = .11, SE = .06, $t(1478) = 1.96$, $p = .051$). The two versions of the *overall + attributes* conditions did not significantly differ from each other using conventional cutoffs for significance, although the effect size was similar as that obtained when comparing the *overall only* and *overall + individual paintings* conditions (b = -.09, SE = .06, $t(1478) = -1.64$, $p = .102$).[6]

This study generalized the results of studies 1-3 to a context in which participants underwent an actual experience. Consumers gave higher overall ratings to a subpar painting gallery when asked to rate both the overall experience along with several features of that experience than when just asked to rate the overall experience. This study also provides some evidence that the

---

[5] In a pretest, we asked a separate group of participants to rate the paintings on a 7-point scale. The means and standard deviations for each painting are as follows: Painting 1, $M = 4.13$, $SD = 1.76$; Painting 2, $M = 2.14$, $SD = 1.54$; Painting 3, $M = 4.16$, $SD = 1.80$; Painting 4, $M = 4.15$, $SD = 1.64$.

[6] These t-values were taken from two pre-registered regressions. First, we regressed the overall rating on an indicator for the *overall + individual paintings* condition and an indicator for the *overall + painting attributes* condition. Then we regressed the overall rating on an indicator for the *overall + individual paintings* condition and an indicator for the *overall only condition*.

effect emerges regardless of whether consumers rate multiple attributes of the experience or each component of the experience (i.e., each painting in a gallery).

STUDY 5

In studies 1-4, we found that consumers rate subpar experiences more positively when they rate specific aspects of those experiences. We hypothesize that this effect emerges because consumers are less likely to incorporate negative aspects of their experience into their overall evaluation when they are specifically invited to voice those aspects. In study 5, we examine whether this effect is specific to asking consumers to provide multiple ratings of specific aspects of their experience or whether it also emerges when consumers are invited to give more open-ended feedback in a text box. On the one hand, providing a text box does give consumers the opportunity to provide specific feedback about negative aspects of their experience. On the other hand, consumers may feel that voicing specific concerns in writing does not supplant the need to also express those concerns more quantitatively, in the form of a lower rating.

Method

Participants. We recruited online participants from Amazon's Mechanical Turk (MTurk) and paid them $0.30. Our pre-registration specified that we would collect data from 1,500 participants and include only the first response in the event of a duplicate IP address or MTurk ID. We removed the responses of 20 duplicate participants, as well as 163 observations from 86

participants who failed to complete the dependent variable. We ended up with 1,492 participants and 2,976 observations (mean age = 37.55, 48.66% female).

Procedure. As in study 2, participants read two scenarios in which they imagined staying at an Airbnb. One scenario described a subpar experience and the other described a good experience, with the order of the scenarios randomized across participants. Between subjects, participants were randomly assigned to either the *overall only* condition, the *overall + attributes* condition, or the *overall + text box* condition. The *overall only* and *overall + attributes* conditions were identical to study 2, as were the scenarios presented to participants. Participants in the *overall + text box* condition provided an overall rating and were also shown a text box and asked, "What review would you provide?" Participants had to provide a written review of at least 15 characters in order to proceed with the survey. At the end of the survey, participants provided their age and gender.

Results and Discussion

Before describing the results, we should note that we observed differential attrition in this study. Participants were more likely to spontaneously drop out of the *overall + text box* condition (8.24%) than the *overall + attributes* condition (4.91%; b = .03, SE = .02, $t(1587) = 2.22$, $p = .027$), and the *overall only* condition (2.08%; b = .06, SE = .01, $t(1587) = 4.65$, $p < .001$). They were also more likely to drop out of the *overall + attributes* condition than the *overall only* condition (b = .03, SE = .01, $t(1587) = 2.55$, $p = .011$).[7] Differential attrition is never desirable, and it can in

---

[7] These *p*-values are from OLS regressions. Logit regressions yielded nearly identical *p*-values.

many cases undermine the validity of random assignment (Simmons and Nelson 2020; Zhou and Fishbach 2016). In web appendix H, we correct for differential attrition in the most conservative way possible, by replacing all excess missing ratings with either maximally positive ratings (i.e., 5s) or maximally negative ratings (i.e., 1s), depending on which one would most effectively undermine our observed finding. As noted below, doing this does alter the significance of some of our findings, but it does not alter our top-line conclusions.[8]

Turning to our main analysis, we regressed participants' overall rating of the Airbnb stay on a contrast-coded indicator for the *overall + attributes* condition (coded as 2/3 for the *overall + attributes* condition and -1/3 for the two other conditions), a contrast-coded indicator for the *overall + text box* condition (coded as 2/3 for the *overall + text box* condition and -1/3 for the two other conditions), a contrast-coded indicator for whether the scenario was subpar (coded as 0.5 for subpar scenarios and -0.5 for good scenarios), the interaction of the *overall + attributes* condition indicator and the subpar scenario indicator, the interaction of the *overall + text box* condition and the subpar scenario indicator, and fixed effects for scenario. We clustered standard errors by participant. Mean overall ratings by condition are displayed in figure 4.[9]

First, we again observed a significant interaction between the *overall + attributes* condition and whether the scenario was subpar ($b = .21$, SE $= .09$, $t(1491) = 2.34$, $p = .019$). As in studies 1-4, participants provided more favorable overall ratings of a subpar experience when they rated it on multiple dimensions than when they simply provided an overall rating ($b = .18$, SE $= .07$, $t(1482) = 2.60$, $p = .009$), though this effect becomes marginal ($p = .068$) when we use a maximally

---

[8] We also found differential attrition in studies 6, 7, and 8; web appendix H shows that our results are robust to our conservative correction procedure for these studies.

[9] In studies 5-7, for ease of exposition, the analytic approach described in the text deviates very slightly from the analytic approach that we pre-registered. In web appendix I, we show that all results are nearly identical – and none of our conclusions are altered – when we run our pre-registered analyses.

conservative approach to correct for differential attrition (see web appendix H). This effect did not emerge when the experience was good (b = -.02, SE = .05, $t(1484)$ = -0.49, p = .625).

Second, although we observed a significant interaction between the *overall + text box* condition and whether the scenario was subpar (b = -.30, SE = .09, $t(1491)$ = -3.51, p < .001), it was opposite in sign to our expectations. When the experience was subpar, providing a text box actually *lowered* participants' overall evaluations, relative to both the *overall only condition* (b = -.25, SE = .07, $t(1482)$ = -3.47, p < .001), and the *overall + attributes* condition (b = -.43, SE = .07, $t(1482)$ = -6.01, p < .001), though the *overall only* vs. *overall + text box* condition difference may be due to differential attrition (see web appendix H). When the experience was good, providing a text box did not influence participants' overall evaluations (ts < 1.66, ps > .097).

FIGURE 4

STUDY 5 RESULTS

NOTE. - Error bars depict +/- 1 standard error.
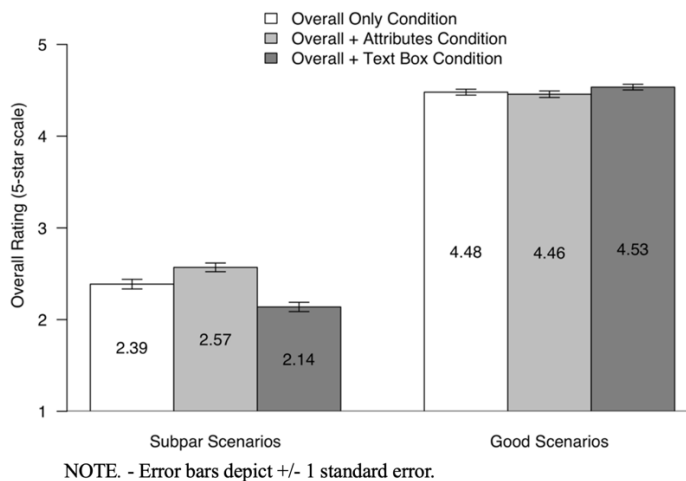
These results suggest that although asking consumers to rate multiple aspects of a subpar experience increases their overall evaluations of that experience, allowing them to describe their

experience in a text box does not. We can only speculate as to why this happens. One possibility is that consumers' inclination to incorporate negative aspects in their overall rating is only reduced when they have an opportunity to provide feedback of a similar form (i.e., another rating), but not of a dissimilar form (i.e., an open-ended description of their experience). It is also possible that consumers simply value providing negative feedback in the form of ratings more than they value providing open-ended negative feedback, perhaps because they believe that open-ended feedback is more likely to be ignored and/or less likely to be influential. Whatever the reason, our results suggest that our effect may be specific to asking participants to actually rate various aspects of their subpar experiences.

WHY DOES RATING SPECIFIC ASPECTS OF A SUBPAR EXPERIENCE INCREASE OVERALL EVALUATIONS OF THAT EXPERIENCE?

In studies 1-5, we consistently found that consumers provide more positive overall ratings of a subpar experience when they are invited to rate specific aspects of that experience. Having established the robustness of this effect, we now turn to understanding why it happens. In studies 6-9, we set out to distinguish among three potential mechanisms: "positive reminders", "averaging", and "avoiding negative redundancy."

According to the positive reminders account, asking consumers to rate specific aspects of subpar experiences increases their overall evaluations of those experiences because it reminds them that their experiences were also comprised of positive aspects. For example, although a restaurant consumer who experiences bad food and good service may be naturally inclined to focus

on the bad food when providing her overall rating, being asked to rate both the food and the service may remind her that the service was actually just fine, leading her to increase her overall rating.

An averaging account, of the type put forth by Schneider et al. (2020), is quite similar. By this account, consumers' overall ratings reflect an average of the attribute ratings that they are asked to provide. Thus, if you ask consumers who have subpar experiences to rate some positive attributes alongside the (already salient) negative ones, their average attribute rating may exceed what their rating was otherwise going to be. If consumers subsequently rely on that average when generating their overall ratings, then their overall ratings will become more positive.

Both of these accounts make a straightforward prediction: asking consumers to rate only *positive* aspects of an experience should *increase* consumers' overall ratings, while asking consumers to rate only *negative* aspects of an experience should *decrease* consumers' overall ratings.

But the avoiding negative redundancy account makes the opposite prediction. By this account, rating multiple aspects of a subpar experience increases consumers' overall ratings precisely because consumers avoid incorporating negative information into both the overall rating and the attribute ratings. In other words, consumers want to be truthful, but they also err on the side of being charitable. If they are given the chance to be truthful about a bad aspect by providing a specific attribute rating, then they do not have to do so (as much) when providing their general, overall rating. Thus, whereas the positive reminders and averaging accounts predict that you can most effectively increase consumers' overall ratings of subpar experiences by having them rate the *positive* aspects of those experiences, the avoiding negative redundancy account predicts that you can most effectively increase consumers' overall ratings of subpar experiences by having them rate the *negative* aspects of those experiences. We test this in studies 6-9.

STUDIES 6 AND 7

Studies 6 and 7 directly tested these accounts by including an *overall + negative attributes* condition (in study 6) and an *overall + positive attributes* condition (in study 7). In the *overall + negative attributes* condition, participants were asked to provide an overall rating and to rate the most negative attributes of the experience, while in the *overall + positive attributes* condition, participants were asked to provide an overall rating and to rate the most positive attributes of the experience. The positive reminder and averaging mechanisms predict that consumers would rate the overall experience most positively in the *overall + positive attributes* condition, followed by the *overall + attributes* condition, and then the *overall only* and *overall + negative attributes* conditions. The avoiding negative redundancy mechanism predicts that consumers would rate the overall experience more positively in the conditions that invite consumers to specifically rate the negative aspects of their experience (i.e., the *overall + negative attributes* and *overall + attributes* conditions) than in the conditions that do not allow consumers to rate the negative aspects of their experience (i.e., the *overall only* and *overall + positive attributes* conditions).

Method

Participants. For both studies, we recruited online participants from Amazon's Mechanical Turk (MTurk) and paid $0.30. Our pre-registrations specified that we would collect data from 1,500 participants and include only the first response in the event of a duplicate IP address or MTurk ID. For study 6, we removed the responses of 10 duplicate participants, as well as 128

observations from 66 participants who failed to complete the dependent variable. Our final sample consisted of 1,499 participants and 2,994 observations (mean age = 36.77, 48.30% female). For study 7, we removed the responses of 20 duplicate participants, as well as 95 observations from 51 participants who failed to complete the dependent variable. Our final sample consisted of 1,490 participants and 2,973 observations (mean age = 35.32, 50.87% female).

Procedure. Both studies used the same scenarios and conditions as study 2, and simply added an additional rating condition. Thus, after reading instructions, participants read and rated a good and subpar Airbnb scenario on separate pages. After doing so, participants provided their age and gender.

Participants in study 6 were randomly assigned to one of three rating conditions: the *overall only* condition, the *overall + attributes* condition, or the *overall + negative attributes* condition. The *overall only* and *overall + attributes* conditions were identical to study 2. In the *overall + negative attributes* condition, participants rated the overall experience and, for each scenario, the two attributes that were rated most negatively in previous studies.[10]

Participants in study 7 were also randomly assigned to one of three rating conditions: the *overall only condition*, the *overall + attributes* condition, or the *overall + positive attributes* condition. Again, the *overall only* and *overall + attributes* conditions were identical to study 2. In the *overall + positive attributes* condition, participants rated the overall experience and, for each scenario, the two attributes that were rated most positively in previous studies.

Results and Discussion

---

[10] It is worth noting, however, that within the good experiences condition, even the most negatively rated attributes were not rated very negatively (Ms > 4.36 out of 5 stars).

Our *overall + negative attributes* and *overall + positive attributes* manipulations were successful. Participants' average ratings of specific features of the experience were more negative in the *overall + negative attributes* condition ($M = 2.11$, SD $= 1.01$) than in the *overall + attributes* condition in study 6 ($M = 2.55$, SD $= 0.85$; d $= -.46$, $t(987) = -7.29$, $p < .001$), and more positive in the *overall + positive attributes* condition ($M = 3.02$, SD $= 1.05$) than in the *overall + attributes* condition in study 7 ($M = 2.68$, SD $= 0.90$; d $= .35$, $t(992) = 5.47$, $p < .001$).

Figure 5 displays the mean overall ratings by condition. We regressed the overall rating on a contrast-coded indicator for the *overall + attributes* condition (coded as 2/3 for the *overall + attributes* condition and -1/3 for the two other conditions), a contrast-coded indicator for either the *overall + negative attributes* condition (study 6) or *overall + positive attributes* condition (study 7; coded as 2/3 for the specified condition and -1/3 for the two other conditions), a contrast-coded indicator for whether the scenario was subpar (coded as 0.5 for subpar scenarios and -0.5 for good scenarios), the interaction of the *overall + attributes* condition indicator and the subpar scenario indicator, the interaction of the *overall + positive* or *overall + negative attributes* condition and the subpar scenario indicator, and fixed effects for scenario. We clustered standard errors by participant.

In both studies, we once again found a significant interaction between the *overall + attributes* condition and whether the experience was subpar (study 6: $b = .32$, SE $= .08$, $t(1498) = 4.19$, $p < .001$; study 7: $b = .26$, SE $= .08$, $t(1489) = 3.20$, $p = .001$). Replicating previous results, participants in the *overall + attributes* condition gave significantly higher overall ratings to subpar experiences than those in the *overall only* condition (study 6: $b = .36$, SE $= .06$, $t(1493) = 5.87$, $p < .001$; study 7: $b = .24$, SE $= .07$, $t(1480) = 3.57$, $p < .001$), and there was no significant difference

between these two conditions for good scenarios (study 6: $b = .04$, SE $= .04$, $t(1,491) = 1.00$, p $= .319$; study 7: $b = -.03$, SE $= .05$, $t(1483) = -.61$, $p = .542$).

In study 6, we also observed a significant interaction between the *overall + negative attributes* condition and whether the experience was subpar ($b = .24$, SE $= .08$, $t(1498) = 3.16$, p $= .002$). Unpacking this interaction, we found that participants in the *overall + negative attributes* condition rated the overall experience significantly more positively than those in the *overall only condition* ($b = .26$, SE $= .06$, $t(1493) = 4.29$, $p < .001$). There was no significant difference between the *overall + attributes* condition and the *overall + negative attributes* condition ($b = .10$, SE $= .06$, $t(1493) = 1.57$, $p = .116$). These results are consistent with the "avoiding negative redundancy" account and inconsistent with the "positive reminder" and "averaging" accounts, both of which would have expected rating negative aspects to decrease rather than increase consumers' overall ratings of subpar experiences.

The results of study 7 were also consistent with the avoiding negative redundancy account and inconsistent with the positive reminders and averaging accounts. For subpar experiences, participants gave significantly higher overall ratings in the *overall + attributes* condition, which invited them to rate both positive *and* negative aspects of the experiences, than in the *overall + positive attributes* condition ($b = .17$, SE $= .07$, $t(1480) = 2.63$, $p = .009$). Moreover, overall ratings of the subpar experience in the *overall + positive attributes* condition did not differ from those in the *overall only* condition ($b = -.06$, SE $= .07$, $t(1480) = -0.93$, $p = .351$), in direct opposition to what the positive reminder and averaging accounts would predict. There was a marginally significant interaction between the *overall + positive attributes* condition and whether the experience was subpar ($b = .16$, SE $= .08$, $t(1489) = 1.87$, $p = .062$), but this was primarily driven by a decrease in overall evaluations of good experiences in the *overall + positive attributes*

condition. This result was unexpected and it does not seem to replicate, so we do not discuss it further.

Overall, the findings of studies 6 and 7 are consistent with the avoiding negative redundancy account, and inconsistent with the positive reminder and averaging accounts.

FIGURE 5

STUDIES 6 AND 7 RESULTS
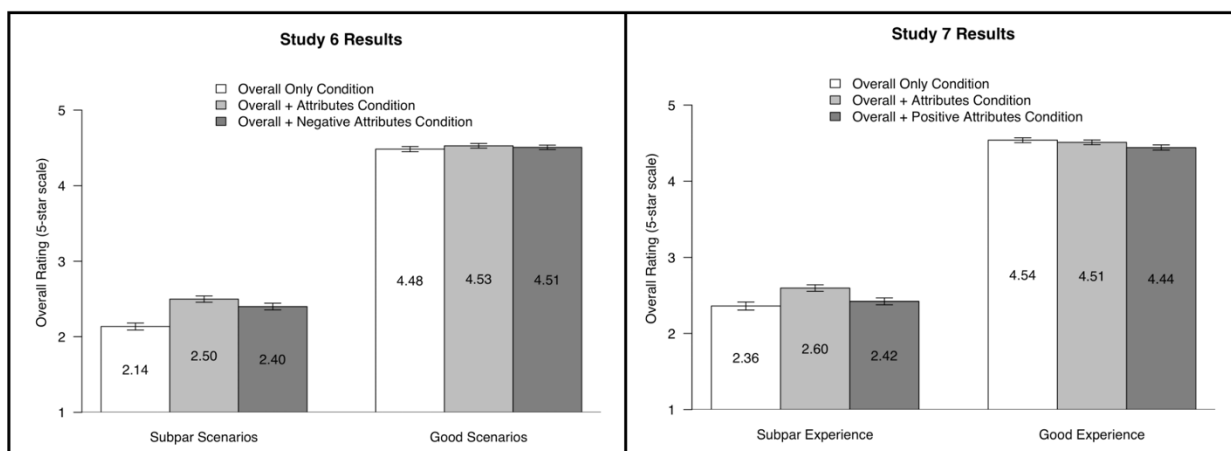
NOTE. - Error bars depict +/- 1 standard error.

STUDIES 8 AND 9

In studies 8 and 9, we sought to directly replicate the results of studies 6 and 7, while focusing specifically on ratings of subpar experiences. We also sought to establish that those results generalize to other contexts and scenarios.

Method

Participants. For both studies, we recruited online participants from Amazon's Mechanical Turk (MTurk). Study 8 paid $0.20 and Study 9 paid $0.55. Our pre-registration for study 8 specified that we would collect data from 2,000 participants, while in study 9 we specified that we would collect data from 1,600 participants. In study 8, we indicated we would only include the first response in the event of a duplicate IP address or MTurk ID. We removed the responses of 25 duplicate participants, as well as 10 observations from 10 participants who failed to complete the dependent variable; our final sample consisted of 1,985 participants and observations (mean age = 35.91, 52.02% female). In study 9, we indicated we would delete any duplicate responses and include only participants who passed the attention check and submitted their response for payment. We removed the responses of 132 duplicate participants, 39 participants who did not submit their response for payment, and 3,742 observations from 937 participants who failed to complete the dependent variable; our final sample consisted of 1,426 participants and 5,704 observations (mean age = 41.60, 53.79% female).[11]

Procedure. In study 8, participants read and rated one scenario about a subpar restaurant experience, and then provided their age and gender. For stimulus sampling purposes, we created two subpar scenarios, and participants were randomly assigned to one of them. Participants rated their experiences on 5-star scales.

In study 9, participants first completed an attention check question, and then read and rated four subpar scenarios, one from each of the following domains: a commercial flight experience, a dentist experience, a sneakers purchase, and a bookcase purchase. The scenarios were presented

---

[11] Notably, 912 of the 937 participants who failed to complete the dependent variable did so because they failed the attention check. As in other studies, this attention check was administered before random assignment and participants were not allowed to continue with the study if they did not answer it correctly.

in a random order. For stimulus sampling purposes, we created three scenarios in the bookcase purchase domain, and four scenarios in the flight experience, dentist experience, and sneakers purchase domains. Within domain, participants were randomly assigned to one of the scenarios that we created. After reading and rating all four scenarios, participants provided their age and gender. Participants rated their experiences on 10-star scales.

In both studies, participants were randomly assigned to the following conditions: the *overall only* condition, the *overall + attributes* condition, the *overall + negative attributes* condition, or the *overall + positive attributes* condition. As in studies 6 and 7, the *overall + negative attributes* condition included one or two attributes that had been pretested to be particularly bad, while the *overall + positive attributes* condition included one or two attributes that had been pretested to be particularly good. In study 8, participants were randomly assigned to one of these conditions. In study 9, participants were randomly assigned (with replacement) to a different rating condition for each of the four scenarios that they evaluated.

Results and Discussion

Our negative and positive attribute manipulations were once again successful. In study 8, participants' average ratings of specific features of the experience were more negative in the *overall + negative attributes* condition ($M = 1.83$, SD = 1.16) than in the *overall + attributes* condition ($M = 2.43$, SD = 0.89; d = -.58, $t(995) = -9.13$, $p < .001$), and more positive in the *overall + positive attributes* condition ($M = 3.14$, SD = 1.21) than in the *overall + attributes* condition (d = .67, $t(992) = 10.49$, $p < .001$). This was true in study 9 as well: Participants' average ratings of specific features of the experience were more negative in the *overall + negative attributes*

condition (M = 2.41, SD = 1.53) than in the *overall + attributes* condition (M = 4.35, SD = 1.52;

d = -1.28, t(2871) = -34.16, p < .001), and more positive in the *overall + positive attributes*

condition (M = 6.23, SD = 2.35) than in the *overall + attributes condition* (d = .95, t(2839) = 25.26,

p < .001). This pattern held across all products tested in study 9 (all *p*s < .001).

FIGURE 6

STUDIES 8 AND 9 RESULTS
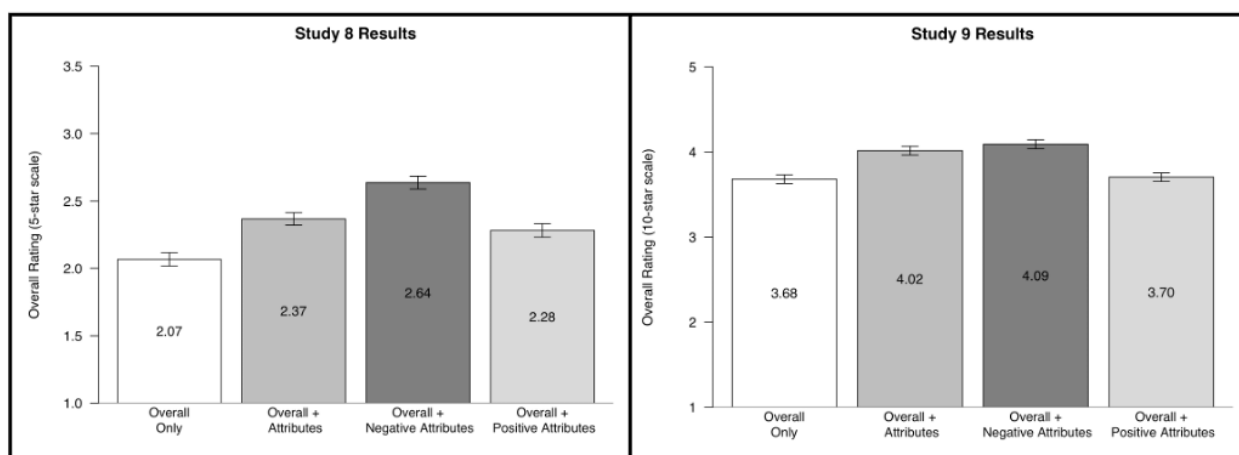
NOTE. - Error bars depict +/- 1 standard error.

Figure 6 displays the mean overall ratings by condition. To compare the other rating

conditions to the *overall only* condition, we regressed the overall rating on an indicator for the

overall + attributes condition, an indicator for the *overall + positive attributes* condition, an

indicator for the *overall + negative attributes* condition, and fixed effects for scenario. All

indicators were dummy coded (e.g., 1 or 0). We once again found that participants in the *overall

+ attributes* condition rated the overall experience significantly higher than those in the *overall

only condition* (study 8: b = .30, SE = .07, t(1980) = 4.47, p < .001; study 9: b = .33, SE = .07,

t(1425) = 4.95, p < .001). Moreover, consistent with the avoiding negative redundancy account,

participants in the *overall + negative attributes* condition rated the overall experience significantly higher than those in the *overall only condition* (study 8: $b = .57$, SE $= .07$, $t(1980) = 8.44$, $p < .001$; study 9: $b = .37$, SE $= .07$, $t(1425) = 5.43$, $p < .001$). Although participants in the *overall + positive attributes* condition of study 8 rated the overall experience significantly higher than those in the *overall only condition* ($b = .22$, SE $= .07$, $t(1980) = 3.19$, $p = .001$), this result did not replicate in study 9 ($b = .03$, SE $= .07$, $t(1425) = 0.40$, $p = .693$).

To compare the other rating conditions to the *overall + attributes* condition, we re-ran our regression, this time including an indicator for the *overall only* condition in place of the indicator for the *overall + attributes* condition. Participants in the *overall + negative attributes* condition gave a significantly *higher* overall rating than those in the *overall + attributes* condition in study 8 ($b = .27$, SE $= .07$, $t(1980) = 3.99$, $p < .001$). In study 9 there was no significant difference between the *overall + negative attributes* and *overall + attributes* condition ($b = .04$, SE $= .07$, $t(1425) = 0.62$, $p = .541$). Both of these results contradict the positive reminders and averaging accounts, which predict that those in the *overall + negative attributes* condition will, by virtue of the attention drawn to the negative attributes of the experience, rate their overall experience significantly *lower* than those in the *overall + attributes* condition. Finally, although there was no significant difference between the *overall + positive attributes* condition and the *overall + attributes* condition in study 8 ($b = -.08$, SE $= .07$, $t(1980) = -1.26$, $p = .207$), those in the *overall + positive attributes* condition of study 9 rated the overall experience lower than those in the *overall + attributes* condition ($b = -.30$, SE $= .07$, $t(1425) = -4.59$, $p < .001$). Once again, this is consistent with the avoiding negative redundancy account and inconsistent with the positive reminders and averaging accounts.

Altogether, the results of studies 6-9 strongly suggest that asking consumers to rate negative, but not positive, aspects of an experience increases their overall ratings of that experience.

STUDY 10

Studies 1-9 found that consumers give higher overall ratings when they rate both the overall experience and several attributes of an experience, compared to rating the overall experience alone. In study 10, we tried to see which overall rating is more accurate: Are consumers giving too much weight to the negative attributes in the *overall only* condition or are they giving too little weight to the negative attributes in the *overall + attributes* condition? We asked participants to imagine a subpar restaurant experience and we randomly assigned them to either the *overall only* or *overall + attributes* condition. Then, on a separate page, we asked participants to complete two measures designed to assess their behavioral intentions toward the restaurant (e.g., how likely they would be to return to the restaurant). A more accurate overall rating should better predict these behavioral intentions; thus, we examined whether the correlation between participants' overall ratings and these behavioral intention measures differed by condition.

Method

Participants. We recruited online participants from Amazon's Mechanical Turk (MTurk) and paid them $0.40. Our pre-registration specified that we would collect data from 2,000 participants, exclude any duplicate participants, and exclude any participants who failed our simple

attention check question. We removed the responses of six duplicate participants, 18 participants who failed to complete the dependent variable, 14 participants who answered our attention check question incorrectly, and six participants whose IDs were not submitted for payment. We ended up with 1,991 participants and observations (mean age = 40.58, 55.14% female).

Procedure. Participants read one scenario in which they imagined going to a restaurant; we used the same two subpar restaurant experience scenarios as in studies 1 and 8, and participants were randomly assigned to one of them. Between subjects, participants were randomly assigned to either the *overall only* condition or the *overall + attributes* condition. The *overall only* and *overall + attributes* conditions were identical to studies 1 and 8. After rating the restaurant, participants moved on to a separate page where they answered the following two behavioral intention questions ($r = .84$): "How likely would you be to go to this restaurant?" (1 = "Not at all likely," 9 = "Extremely likely"); "How likely would you be to recommend a friend to go to this restaurant?" (1 = "Would definitely not recommend," 9 = "Would definitely recommend"). At the end of the survey, participants provided their age and gender.

Results and Discussion

First, we conducted a t-test to examine whether the overall rating differed by condition. We found that participants in the *overall + attributes* condition rated the experience significantly more positively ($M = 3.43$, SD = 1.70) than those in the *overall only* condition ($M = 2.70$, SD = 1.71; d = .44, $t(1989) = 9.87$, $p < .001$).

Next, we tested whether responses to the average of the two behavioral intention measures differed by condition. They did not, as these behavioral intention ratings were nearly identical in the *overall only* condition (*M* = 1.71, SD = 1.13) and the *overall + attributes* condition (*M* = 1.70, SD = 1.17; d = .00, *t*(1989) = .09, *p* = .93). Thus, even though participants' overall ratings were different in the different conditions, this difference did not manifest in their behavioral intentions. This suggests that the condition difference in overall ratings does not reflect true differences in beliefs as much as different strategies of responding. And if that is the case, then it raises the question of which condition's ratings are more accurate.

To answer that question, for each condition we computed the correlation between participants' overall ratings and the average of the behavioral intention measures. We found that that correlation was significantly larger in the *overall only* condition (*r* = .67, *t*(996) = 28.82, *p* < .001) than in the *overall + attributes* condition (*r* = .56, *t*(991) = 21.31, *p* < .001; *z* = 4.12, *p* < .001; additional analyses can be found in web appendix L). This suggests that, compared to an *overall only* regime, asking consumers to rate all of the attributes of a subpar experience may artificially increase the ratings they provide while decreasing the predictive value of those ratings. Notably, this is not to say that attribute ratings are unhelpful: having more information (in this case, in the form of attribute ratings) is usually helpful. Instead, these results highlight that eliciting attribute ratings affects the correlation between the overall rating and consumers' behavioral intentions, thus affecting the diagnosticity of those overall ratings.

GENERAL DISCUSSION

In a series of pre-registered experiments, we found that asking consumers to rate specific aspects of their subpar experiences increased how positively they rated those experiences overall. Studies 1 and 2 established the basic effect across multiple domains and scenarios, while showing that it does not hold for "good" experiences, for which no aspects are negative. Study 3 showed that the effect was robust to recalled experiences, while study 4 established that the effect generalizes to ratings of a real (online) experience. Study 5 found that the effect is limited to being asked to *rate* specific features of an experience, rather than being asked to provide open-ended comments about those features. And in study S1, in web appendix A, we found that the effect holds even when consumers are merely told that they are going to rate specific aspects on a subsequent page.

Having shed light on the circumstances under which the effect holds, we conducted studies 6-9 to try to understand why it happens. We considered two categories of mechanisms. One possibility is that rating multiple aspects of subpar experiences reminds consumers that there were also some positive features of those experiences (Earthy et al. 1997). For example, rating the specific attributes of a restaurant experience that featured bad service could remind consumers that the food and ambience were actually quite good. As a consequence, consumers may be more likely to incorporate those positive features into their overall ratings, thereby increasing them. Schneider et al. (2020) recently put forth a similar mechanism when they suggested that consumers' overall ratings reflect an average of their ratings of specific aspects, implying that rating more positive aspects of an experience should increase the positivity of consumers' overall ratings.

Despite the intuitive appeal of, and apparent empirical support for, these mechanisms, our findings support a different account, one that we refer to as avoiding negative redundancy. According to this account, many raters of subpar experiences are conflicted: they want to be

truthful, but they also want to avoid being too negative, preferring to err on the side of positivity. Being asked to provide only a single overall rating forces them to be truthful, to heavily incorporate the negative aspects of their experiences into that rating. But being asked to additionally rate the specific aspects of their experiences allows them to achieve both aims at once: they can rate the experience poorly on the specific negative aspect(s), while erring on the side of positivity when providing their overall rating.

Importantly, these two accounts make different predictions. Whereas the positive reminders account predicts that consumers will provide more positive overall ratings when they are asked to rate the *positive* aspects of their experiences, the avoiding negative redundancy account predicts that consumers will provide more positive overall ratings when they are asked to rate the *negative* aspects of their experiences. Consistent with the latter, in studies 6-9 we found that although rating *negative* aspects of a subpar experience consistently increased consumers' overall ratings of that experience, rating *positive* aspects did not.

The avoiding negative redundancy account also predicts that consumers will be more likely to artificially inflate their overall ratings when they are asked to rate negative aspects of their experiences, and that those ratings would therefore be less truthful. Consistent with this, in study 10 we found that eliciting ratings of specific aspects of their experiences led consumers to generate overall ratings that were less predictive of their behavioral intentions.

Theoretical and Practical Implications

Our research is not the first to suggest or show that asking consumers to rate specific attributes of an experience may influence their overall rating of that experience (Chen, Hong and

Liu 2018; Earthy et al. 1997; Schneider et al. 2020). But it is the first to identify "avoiding negative redundancy" as a mechanism of the effect, and to thereby identify important boundary conditions of it. We are the first to show that rating specific attributes is more likely to increase consumers' overall ratings when (1) the experience is subpar, and (2) consumers are asked to rate the *negative* features of that experience.

By proposing and showing evidence for this novel mechanism, we contribute to several extant literatures. For instance, this work adds to the large literature on consumer ratings by providing a novel explanation for how the addition of attribute ratings alters consumers' overall ratings (Burtch et al. 2018; Chevalier and Mayzlin 2006; Chintagunta et al. 2010; Donati 2022; Goes et al. 2014; He and Bond 2015; Kwark et al. 2014; Simonson 2016; Zhu and Zhang 2010). It also adds to work on attitude assessment (Argyriou and Melewar 2011; Fazio et al. 1984; Lutz 1991; Van Overwalle and Siebler 2005) by demonstrating how the addition of attribute ratings can change how consumers report their attitudes about products and experiences. And, finally, it enhances our understanding of how to quell negativity bias in this context. Whereas negativity bias may lead consumers to focus more on negative rather than positive aspects of their experiences (Baumeister et al. 2001; Fiske 1981; Rozin and Royzman 2001), affording them the opportunity to specifically rate the negative aspects may license them to focus more on the positive aspects, or at least to express attitudes that more heavily weigh those aspects. Thus, we add to a growing literature exploring moderators of negativity bias in online reviews and ratings (Chen and Lurie 2013; Wu 2013).

Relatedly, it is worth commenting on how our research relates to Gal and Rucker's (2011) seminal work on "response substitution." Gal and Rucker find that consumers who have a subpar experience rate positive attributes more negatively (e.g., the food at a restaurant) if they are not

given the opportunity to rate salient negative aspects of the experience (e.g., the service at a restaurant). Thus, just as the participants in our studies give higher overall ratings when they are given an opportunity to rate the experience on its specific negative feature(s), participants in Gal and Rucker's research give higher ratings to another specific attribute when they are given an opportunity to rate the experience on its specific negative feature(s). In explaining this effect, Gal and Rucker wrote that "people might have a need to explicitly express their attitudes" (p. 186), a contention that resonates with our suggestion that consumers may have a desire to truthfully report the negative aspects of their experience. Once that desire is fulfilled by rating specific negative attributes, consumers may then feel licensed to give more positive overall (or attribute-specific) ratings of the experience.

Although our mechanism is similar to that of Gal and Rucker's (2011), there are important differences between our account and theirs. First, whereas Gal and Rucker's research focuses on ratings (of a specific attribute) that should not incorporate consumers' attitudes toward other attributes (e.g., ratings of the food should not incorporate consumers' attitudes toward the service), our research focuses on ratings (of the overall experience) that *should* incorporate consumers' attitudes toward the attributes. There is no reason why our results need to be the same. Indeed, whereas a restaurant consumer who rates both the good food and the bad service knows that her food rating should not incorporate the service, a consumer who rates both the bad service and her overall experience probably knows that her overall rating *should* incorporate the service. Thus, the normative implication to provide different ratings when both are elicited is present in Gal and Rucker's circumstance, but not in ours.

Second, in Gal and Rucker's (2011) research, restaurant consumers who want to rate the (bad) service but are only asked to rate the (good) food tend to artificially deflate their ratings of

the food. This means that if you want to make consumers' attribute ratings more accurate, you should ask them to rate all of the attributes rather than just one attribute in isolation. In this context it is interesting to consider our study 10, in which we found that the correlation between overall ratings and behavioral intentions was *lower* when participants were asked to rate the specific aspects of their subpar experiences than when they simply provided an overall rating. This suggests that the diagnosticity, or predictive validity, of participants' overall ratings, is *worsened* by having consumers rate all the specific attributes. In other words, it seems that asking consumers to rate the specific aspects of a subpar experience may lead them to artificially inflate their overall ratings; their "true" overall rating seems to be the one made in isolation, not the one made alongside all the attributes. Thus, whereas Gal and Rucker's research suggests that more accurate attribute ratings can be elicited by having consumers rate all of the attributes, our research suggests that more accurate overall ratings can be elicited by having consumers provide just a single overall rating.

Of course, these two conclusions are not at odds, but rather reflect the different foci of these investigations. If you want consumers to accurately rate *attributes* of their experience, then maybe you should ask them to rate them all. But if you want them to accurately rate their *overall experience*, then maybe you should just ask them to do that and nothing else.

A theoretical challenge we face with this research is in trying to reconcile our results with those of Schneider et al. (2020), who in one study (study 2c) found support for an "averaging account": Participants gave higher overall evaluations when they were asked to rate more positive features of an experience. In that study, online participants watched a video in which a number of stick figures were killed by another stick figure. While watching the video, participants were asked to count how many stick figures were killed. They then went on to rate specific aspects of the

video. In the *overall + positive attributes* condition, they rated the video on three positive attributes: Visual Effects, Animation Quality, and Stream Quality. In the *overall + mixed attributes* condition, they rated the video on two positive attributes (Visual Effects and Animation Quality) and one negative attribute (Story). Then, further down the page, they provided an overall rating. The key result, and the one that is contrary to ours, is that participants rated the movie more positively overall after rating three positive features than after rating two positive features and a negative feature.

We believe that the discrepancy between our results and theirs may be resolved as follows. Whereas in our studies the negative aspects of a consumer's experience were quite salient even for participants who were not asked to rate those aspects – for example, they *knew* whether the service was bad or the Airbnb rental was dirty – in this study the negative aspect of the experience – the subpar story – was likely not at all salient to participants who were not asked to rate it. When participants watched the stick figure video, they were probably not thinking to themselves, "This is really missing a good story," because the video wasn't designed to offer a good story. Indeed, there *is* no story, just violent exchanges among stick figures. But once participants were asked about the story, it necessarily became salient to them, and they may have believed that they were supposed to incorporate it into their overall rating (Grice 1975). Because the attributes of a consumption experience, like the service received at a restaurant or the cleanliness of an Airbnb, are usually quite salient to consumers, we expect that real world settings will more often resemble the settings that we investigated, and thus that our results would hold in those settings. But, of course, future research should attempt to verify this.

Altogether, our results also have potentially important practical implications. First, we should note that our results represent, on average, a Cohen's d of .27, or about 0.30 points on a 5-

star rating scale. On the one hand, this is, like many interventions of a similar scope, not a particularly large effect (DellaVigna and Linos 2022). On the other hand, an average increase of 0.3 points means a 30% chance of increasing the left digit of an average rating (e.g., from 2.72 to 3.02), which could have a sizable effect on how consumers respond to those ratings (Fisher et al. 2018; Manning and Sprott 2009; Sokolova, Seenivasan and Thomas 2020; Thomas and Morwitz 2005).

Our results suggest that altering the elicitation method of consumer ratings affects what ratings consumers give, without changing their beliefs about the product or experience. Firms, rating platforms, or marketing managers can adjust the way that they elicit ratings from consumers based on what those entities want to accomplish. If they want to maximize consumers' ratings, or if they simply value the specific feedback that attribute ratings provide, it might make sense for them to ask for attribute ratings in addition to an overall rating. For example, imagine that an airline is designing a survey to collect feedback from customers after their flight. The airline wants positive overall ratings, which they can then highlight in future advertisements. Our work would suggest that this airline can best elicit positive overall ratings if they also elicit ratings of specific attributes, so long as some of those attributes are both negative and already salient to consumers.

But for entities that want to maximize the accuracy or predictive value of consumers' overall ratings, it may instead make sense to elicit just the overall rating and nothing else. For instance, a rating platform like Yelp presumably wants their ratings to be diagnostic of what the actual experience is like so that consumers trust the ratings and use them. Our work would suggest that if the platform's goal is to maximize the diagnosticity of the ratings, it should only ask for the overall rating. Of course, doing this comes at the expense of collecting more information (via attribute ratings) that could be used to conduct additional analyses (e.g., examine weighted

averages of the attributes). This tension highlights the tradeoffs firms face when deciding which rating system to use.

Limitations and Future Directions

Our research does suffer from some limitations, and some important open questions remain. First, it is important to mention that web appendices B-D describe three studies in which we failed to find significant support for our hypothesized effect. Study S2 finds directional but non-significant support and was probably underpowered. Studies S3 and S4 had participants try to recall a past experience (of any quality) and provide ratings of it. In hindsight, these two studies were probably not well suited to investigating our effect, because most recalled experiences were quite positive and our effect only emerges for subpar experiences. Study 3 directly addressed these issues, asking participants to recall specific subpar experiences and finding the same effect as in the hypothetical studies. Web appendices E and G also report successful conceptual replications of our effect (studies S5 and S7), while showing that it generalizes to displaying the overall rating below the attribute ratings and to experiences that have a balanced mix of positive and negative features.

This research leaves some questions unanswered. First, it would be interesting to explore whether this effect is robust to rating people (e.g., politicians, professors, doctors, etc.). On the one hand, consumers might still want to avoid being repetitive with negative information, leading them to give the negative information less weight when they can directly rate it via attribute ratings. On the other hand, there may be some instances in which consumers are okay with being negative twice. For instance, Democrats asked to rate Donald Trump's presidency might be okay with

incorporating negative information twice, leading the effect to attenuate. Of course, an attenuation in this case could also be due to a floor effect if most Democrats give the lowest possible rating. Nevertheless, future research could explore this.

Second, in this paper, we focused on recent experiences that were short in duration (e.g., a recent restaurant experience or Airbnb stay). It remains unclear whether this effect would emerge for experiences that happened a long time ago, or lasted a while in duration. For those kinds of experiences, consumers rely more heavily on their memory to recall what happened. Thus, a positive or negative reminder mechanism is more likely to occur. For example, if a consumer is being asked to rate her college experience, she might generally think of her college experience as good. But if she is asked to rate the dining halls at her university, that might serve as a reminder that the dining halls were quite bad and lead her to give a lower overall rating as a result (i.e., a negative reminder mechanism). Of course, if she had remembered the dining hall quality initially, or if she could not recall the dining hall quality after seeing the attribute rating for this aspect, then the presence of the dining hall attribute would not alter her overall rating. Thus, any effect of including the attribute ratings for such experience will depend on what consumers remember without prompting, and what information is retrieved from memory after viewing the attribute ratings. Future research could explore these ideas further.

DATA COLLECTION STATEMENT

Data from Study 1 was collected on February 18, 2020 on Amazon Mechanical Turk. Data from Study 2 was collected on March 5, 2020 on Amazon Mechanical Turk. Data from Study 3 was collected on January 24, 2022 on Amazon Mechanical Turk. Data from Study 4 was collected on March 9, 2021 on Prolific. Data from Study 5 was collected on March 19, 2020 on Amazon Mechanical Turk. Data from Study 6 was collected on March 26, 2020 on Amazon Mechanical Turk. Data from Study 7 was collected on April 3, 2020 on Amazon Mechanical Turk. Data from Study 8 was collected on April 22, 2020 on Amazon Mechanical Turk. Data from Study 9 was collected between April 20 and April 24, 2021 on Amazon Mechanical Turk. Data from Study 10 was collected between January 26 and 28, 2022 on Amazon Mechanical Turk. Data from Study S1 was collected on March 16, 2020 on Amazon Mechanical Turk. Data from Study S2 was collected on November 19, 2020 on Prolific. Data from Study S3 was collected on December 8, 2020 on Amazon Mechanical Turk. Data from Study S4 was collected between December 16 and 21, 2020 on Amazon Mechanical Turk. Data from Study S5 was collected on September 26, 2022 on Prolific. Data from Study S6 was collected on June 16, 2021 on Prolific. Lastly, data from Study S7 was collected on October 7, 2021 on Prolific. The first author collected and analyzed data from all studies. Our data, code, preregistrations, and materials can be found at https://researchbox.org/196.

REFERENCES

Ajzen, Icek (2001), "Nature and Operation of Attitudes," *Annual Review of Psychology*, 52(1), 27–58.

Argyriou, Evmorfia and T. C. Melewar (2011), "Consumer Attitudes Revisited: A Review of Attitude Theory in Marketing Research," *International Journal of Management Reviews*, 13(4), 431–51.

Asubonteng, Patrick, Karl J. McCleary, and John E. Swan (1996), "SERVQUAL Revisited: A Critical Review of Service Quality," *Journal of Services Marketing*, 10(6), 62–81.

Bahia, Kamilia and Jacques Nantel (2000), "A Reliable and Valid Measurement Scale for the Perceived Service Quality of Banks," *International Journal of Bank Marketing*, 18(2), 84–91.

Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs (2001), "Bad Is Stronger than Good," *Review of General Psychology*, 5(4), 323–70.

Bennett, Roger (1997), "Anger, Catharsis, and Purchasing Behavior Following Aggressive Customer Complaints," *Journal of Consumer Marketing*, 14(2), 156–72.

Berg, Lisbet, Dag Slettemeås, Ingrid Kjørstad, and Thea Grav Rosenberg (2020), "Trust and the Don't-Want-to-Complain Bias in Peer-to-Peer Platform Markets," *International Journal of Consumer Studies*, 44(3), 220–31.

Berger, Jonah (2014), "Word of Mouth and Interpersonal Communication: A Review and

Directions for Future Research," *Journal of Consumer Psychology*, 24(4), 586–607.

Berger, Jonah and Raghuram Iyengar (2013), "Communication Channels and Word of Mouth:

How the Medium Shapes the Message," *Journal of Consumer Research*, 40(3), 567–79.

Bickart, Barbara A. (1993), "Carryover and Backfire Effects in Marketing Research," *Journal of Marketing Research*, 30(1), 52–62.

Bickart, Barbara A., Joan M. Phillips, and Johnny Blair (2006), "The Effects of Discussion and

Question Wording on Self and Proxy Reports of Behavioral Frequencies," *Marketing Letters*, 17(3), 167–80.

Bienstock, Carol C., John T. Mentzer, and Monroe Murphy Bird (1997), "Measuring Physical

Distribution Service Quality," *Journal of the Academy of Marketing Science*, 25(1), 31–44.

Blodgett, Jeffrey G., Kirk L. Wakefield, and James H. Barnes (1995), "The Effects of Customer

Service on Consumer Complaining Behavior," *Journal of Services Marketing*, 9(4), 31–42.

Brown, Stephen W. and Teresa A. Swartz (1989), "A Gap Analysis of Professional Service

Quality," *Journal of Marketing*, 53(2), 92–98.

Burtch, Gordon, Yili Hong, Ravi Bapna, and Vladas Griskevicius (2018), "Stimulating Online

Reviews by Combining Financial Incentives and Social Norms," *Management Science*, 64(5), 2065–82.

Chen, Pei-Yu, Yili Hong, and Ying Liu (2018), "The Value of Multidimensional Rating Systems: Evidence from a Natural Experiment and Randomized Experiments," *Management Science*, 64(10), 4629–47.

Chen, Zoey and Nicholas H. Lurie (2013), "Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth," *Journal of Marketing Research*, 50(4), 463–76.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43(3), 345–54.

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets," *Marketing Science*, 29(5), 944–57.

Chu, Wujin, Minjung Roh, and Kiwan Park (2015), "The Effect of the Dispersion of Review Ratings on Evaluations of Hedonic Versus Utilitarian Products," *International Journal of Electronic Commerce*, 19(2), 95–125.

Chung, Cindy M.Y. and Peter R. Darke (2006), "The Consumer as Advocate: Self-Relevance, Culture, and Word-of-Mouth," *Marketing Letters*, 17(4), 269–79.

Clemons, Eric K., Guodong Gordon Gao, and Lorin M. Hitt (2006), "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems*, 23(2), 149–71.

Coelho, Pedro S. and Susana P. Esteves (2007), "The Choice between a Fivepoint and a Ten-Point Scale in the Framework of Customer Satisfaction Measurement," *International Journal of Market Research*, 49(3), 313–39.

Collier, Joel E. and Carol C. Bienstock (2006), "Measuring Service Quality in E-Retailing," *Journal of Service Research*, 8(3), 260–75.

Cronin Jr, J. Joseph and Steven A. Taylor (1992), "Measuring Service Quality: A Reexamination and Extension," *Journal of Marketing*, 56(3), 55–68.

DellaVigna, Stefano and Elizabeth Linos (2022), "RCTs to Scale: Comprehensive Evidence From Two Nudge Units," *Econometrica*, 90(1), 81–116.

DeMoranville, Carol W. and Carol C. Bienstock (2003), "Question Order Effects in Measuring Service Quality," *International Journal of Research in Marketing*, 20(3), 217–31.

Donati, Dante (2022), "The End of Tourist Traps: A Natural Experiment on the Impact of Tripadvisor on Quality Upgrading."

Earthy, Philippa J., Halliday J.H. MacFie, and Duncan Hedderley (1997), "Effect of Question Order on Sensory Perception and Preference in Central Location Trials," *Journal of Sensory Studies*, 12(3), 215–37.

Fazio, Russell, Tracy Lenn, and Edwin Effrein (1984), "Spontaneous Attitude Formation," *Social Cognition*, 2(3), 217–34.

Fisher, Matthew, George E. Newman, and Ravi Dhar (2018), "Seeing Stars: How the Binary Bias
    Distorts the Interpretation of Customer Ratings," *Journal of Consumer Research*, 45(3),
    471–89.

Fiske, Susan T. (1981), "Attention and Weight in Person Perception: The Impact of Negative and
    Extreme Behavior.," *Journal of Personality and Social Psychology*, 38(6), 889–906.

Gal, David and Derek D. Rucker (2011), "Answering the Unasked Question: Response
    Substitution in Consumer Surveys," *Journal of Marketing Research*, 48(1), 185–95.

Goes, Paulo B., Mingfeng Lin, and Ching-Man Au Yeung (2014), "'Popularity Effect' in User-
    Generated Content: Evidence from Online Product Reviews," *Information Systems
    Research*, 25(2), 222–38.

Grice, Herbert P. (1975), "Logic and Conversation," in *Speech Acts*, New York: Academic Press,
    41–58.

Halstead, Diane and Thomas J. Page Jr. (1992), "The Effects of Satisfaction and Complaining
    Behavior on Consumer Repurchase Intentions," *The Journal of Consumer Satisfaction,
    Dissatisfaction and Complaining Behavior*, 5, 1–11.

He, Stephen X. and Samuel D. Bond (2015), "Why Is the Crowd Divided? Attribution for
    Dispersion in Online Word of Mouth," *Journal of Consumer Research*, 41(6), 1509–27.

Hennig-Thurau, Thorsten, Kevin P. Gwinner, Gianfranco Walsh, and Dwayne D. Gremler (2004),
    "Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates

Consumers to Articulate Themselves on the Internet?," *Journal of Interactive Marketing*, 18(1), 38–52.

Hunt, H. Keith (1991), "Consumer Satisfaction, Dissatisfaction, and Complaining Behavior," *Journal of Social Issues*, 47(1), 107–17.

Jacobsen, Grant D. (2015), "Consumers, Experts, and Online Product Evaluations: Evidence from the Brewing Industry," *Journal of Public Economics*, 126, 114–23.

Jiang, Yabing and Hong Guo (2015), "Design of Consumer Review Systems and Product Pricing," *Information Systems Research*, 26(4), 714–30.

Kang, Gi-Du and Jeffrey James (2004), "Service Quality Dimensions: An Examination of Grönroos's Service Quality Model," *Managing Service Quality: An International Journal*, 14(4), 266–77.

Khare, Adwait, Lauren I. Labrecque, and Anthony K. Asare (2011), "The Assimilative and Contrastive Effects of Word-of-Mouth Volume: An Experimental Examination of Online Consumer Ratings," *Journal of Retailing*, 87(1), 111–26.

Krosnick, Jon A. and Duane F. Alwin (1987), "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement," *Public Opinion Quarterly*, 51(2), 201–19.

Kwark, Young, Jianqing Chen, and Srinivasan Raghunathan (2014), "Online Product Reviews: Implications for Retailers and Competing Manufacturers," *Information Systems Research*, 25(1), 93–110.

Landrum, Hollis, Victor R. Prybutok, Leon A. Kappelman, and Xiaoni Zhang (2008), "SERVCESS: A Parsimonious Instrument to Measure Service Quality and Information System Success," *Quality Management Journal*, 15(3), 17–25.

de Langhe, Bart, Philip M. Fernbach, and Donald R. Lichtenstein (2016), "Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings," *Journal of Consumer Research*, 42(6), 817–33.

Lehtinen, Uolevi and Jarmo R. Lehtinen (1991), "Two Approaches to Service Quality Dimensions," *The Service Industries Journal*, 11(3), 287–303.

Li, Mengxiang, Liqiang Huang, Chuan-Hoo Tan, and Kwok-Kee Wei (2013), "Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features," *International Journal of Electronic Commerce*, 17(4), 101–36.

Lutz, Richard J. (1991), "The Role of Attitude Theory in Marketing," *Perspectives in Consumer Behavior*.

Lytle, Richard S., Peter W. Hom, and Michael P. Mokwa (1998), "SERV∗OR: A Managerial Measure of Organizational Service-Orientation," *Journal of Retailing*, 74(4), 455–89.

Manning, Kenneth C. and David E. Sprott (2009), "Price Endings, Left-Digit Effects, and Choice," *Journal of Consumer Research*, 36(2), 328–35.

Melumad, Shiri, J. Jeffrey Inman, and Michel Tuan Pham (2019), "Selectively Emotional: How Smartphone Use Changes User-Generated Content," *Journal of Marketing Research*, 56(2), 259–75.

Messinger, Paul R., Jin Li, Eleni Stroulia, Dennis Galletta, Xin Ge, and Sungchul Choi (2009),
"Seven Challenges to Combining Human and Automated Service," *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration*, 26(4), 267–85.

Moe, Wendy W. and David A. Schweidel (2012), "Online Product Opinions: Incidence, Evaluation, and Evolution," *Marketing Science*, 31(3), 372–86.

Mudambi, Susan M. and David Schuff (2010), "Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.Com," *MIS Quarterly*, 34(1), 185–200.

Nowlis, Stephen M., Barbara E. Kahn, and Ravi Dhar (2002), "Coping with Ambivalence: The Effect of Removing a Neutral Option on Consumer Attitude and Preference Judgments," *Journal of Consumer Research*, 29(3), 319–34.

Nyer, Prashanth U. (1999), "Cathartic Complaining as a Means of Reducing Consumer Dissatisfaction," *The Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 12.

——— (2000), "An Investigation into Whether Complaining Can Cause Increased Consumer Satisfaction," *Journal of Consumer Marketing*, 17, 9–19.

Pan, Yue and Jason Q. Zhang (2011), "Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews," *Journal of Retailing*, 87(4), 598–612.

Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1988), "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *1988*, 64(1), 12–40.

Parasuraman, A., Valarie A. Zeithaml, and Arvind Malhotra (2005), "ES-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality," *Journal of Service Research*, 7(3), 213–33.

Peterson, Robert A. and William R. Wilson (1992), "Measuring Customer Satisfaction: Fact and Artifact," *Journal of the Academy of Marketing Science*, 20(1), 61–71.

Pitt, Leyland F., Richard T. Watson, and C. Bruce Kavan (1995), "Service Quality: A Measure of Information Systems Effectiveness," *MIS Quarterly*, 19(2), 173–87.

Rozin, Paul and Edward B. Royzman (2001), "Negativity Bias, Negativity Dominance, and Contagion," *Personality and Social Psychology Review*, 5(4), 296–320.

Santos, Jessica (2003), "E-service Quality: A Model of Virtual Service Quality Dimensions," *Managing Service Quality: An International Journal*, 13(3), 233–46.

Schneider, Christoph, Markus Weinmann, Peter Mohr, and Jan vom Brocke (2020), "When the Stars Shine Too Bright: The Influence of Multidimensional Ratings on Online Consumer Ratings," *Management Science*, 67(6), 3871–98.

Schwarz, Norbert (1999), "Self-Reports: How the Questions Shape the Answers.," *American Psychologist*, 54(2), 93–105.

Schwarz, Norbert and Herbert Bless (1992), "Assimilation and Contrast Effects in Attitude Measurement: An Inclusion/Exclusion Model," *ACR North American Advances*.

Schwarz, Norbert and Fritz Strack (1991), "Context Effects in Attitude Surveys: Applying Cognitive Theory to Social Research," *European Review of Social Psychology*, 2(1), 31–50.

Schwarz, Norbert, Fritz Strack, and Hans-Peter Mai (1991), "Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis," *Public Opinion Quarterly*, 55(1), 3–23.

Seth, Nitin, S.G. Deshmukh, and Prem Vrat (2005), "Service Quality Models: A Review," *International Journal of Quality & Reliability Management*, 22(9), 913–49.

Simmons, Carolyn J., Barbara A. Bickart, and John G. Lynch Jr. (1993), "Capturing and Creating Public Opinion in Survey Research," *Journal of Consumer Research*, 20(2), 316–29.

Simmons, Joseph and Leif Nelson (2020), "[89] Data Replicada #6: The Problem of (Weird) Differential Attrition," *Data Colada*, http://datacolada.org/89.

Simonson, Itamar (2016), "Imperfect Progress: An Objective Quality Assessment of the Role of User Reviews in Consumer Decision Making, A Commentary on de Langhe, Fernbach, and Lichtenstein," *Journal of Consumer Research*, 42(6), 840–45.

Sivadas, Eugene and Jamie L. Baker-Prewitt (2000), "An Examination of the Relationship between Service Quality, Customer Satisfaction, and Store Loyalty," *International Journal of Retail & Distribution Management*, 28(2), 73–82.

Sokolova, Tatiana, Satheesh Seenivasan, and Manoj Thomas (2020), "The Left-Digit Bias: When and Why Are Consumers Penny Wise and Pound Foolish?," *Journal of Marketing Research*, 57(4), 771–88.

Sridhar, Shrihari and Raji Srinivasan (2012), "Social Influence Effects in Online Product Ratings," *Journal of Marketing*, 76(5), 70–88.

Sundaram, D. S., Kaushik Mitra, and Cynthia Webster (1998), "Word-Of-Mouth Communications: A Motivational Analysis," *ACR North American Advances*, NA-25, https://www.acrwebsite.org/volumes/8208/volumes/v25/NA-25/full.

Thomas, Manoj and Vicki Morwitz (2005), "Penny Wise and Pound Foolish: The Left-Digit Effect in Price Cognition," *Journal of Consumer Research*, 32(1), 54–64.

Tormala, Zakary L. and Derek D. Rucker (2007), "Attitude Certainty: A Review of Past Findings and Emerging Perspectives," *Social and Personality Psychology Compass*, 1(1), 469–92.

——— (2018), "Attitude Certainty: Antecedents, Consequences, and New Directions," *Consumer Psychology Review*, 1(1), 72–89.

Tourangeau, Roger and Kenneth A. Rasinski (1988), "Cognitive Processes Underlying Context Effects in Attitude Measurement.," *Psychological Bulletin*, 103(3), 299.

Tsekouras, Dimitrios (2017), "The Effect of Rating Scale Design on Extreme Response Tendency in Consumer Product Ratings," *International Journal of Electronic Commerce*, 21(2), 270–96.

Urden, Linda D. (2002), "Patient Satisfaction Measurement: Current Issues and Implications," *Professional Case Management*, 7(5), 194–200.

Van Overwalle, Frank and Frank Siebler (2005), "A Connectionist Model of Attitude Formation and Change," *Personality and Social Psychology Review*, 9(3), 231–74.

Weijters, Bert and Hans Baumgartner (2012), "Misresponse to Reversed and Negated Items in Surveys: A Review," *Journal of Marketing Research*, 49(5), 737–47.

Weijters, Bert, Elke Cabooter, and Niels Schillewaert (2010), "The Effect of Rating Scale Format on Response Styles: The Number of Response Categories and Response Category Labels," *International Journal of Research in Marketing*, 27(3), 236–47.

West, Patricia M. and Susan M. Broniarczyk (1998), "Integrating Multiple Opinions: The Role of Aspiration Level on Consumer Response to Critic Consensus," *Journal of Consumer Research*, 25(1), 38–51.

Wu, Philip Fei (2013), "In Search of Negativity Bias: An Empirical Study of Perceived Helpfulness of Online Reviews," *Psychology & Marketing*, 30(11), 971–84.

Zhou, Haotian and Ayelet Fishbach (2016), "The Pitfall of Experimenting on the Web: How Unattended Selective Attrition Leads to Surprising (yet False) Research Conclusions," *Journal of Personality and Social Psychology*, 111(4), 493–505.

Zhu, Feng and Xiaoquan Zhang (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing*, 74(2), 133–48.