American Psychological Association

Journal of Experimental Psychology: Applied

1076-898X/18/$12.00

2018, Vol. 24, No. 3, 296–305
http://dx.doi.org/10.1037/xap0000164

The Adding-and-Averaging Effect in Bundles of Information: Preference Reversals Across Joint and Separate Evaluation

Kimberlee Weaver
University of Haifa

Stephen M. Garcia
University of Michigan

When does adding mildly favorable information (e.g., job experience at the Railway Credit Union) alongside highly favorable (e.g., job experience at Goldman Sachs) *increase* versus *decrease* evaluations of a bundle of information like a resume or product bundle? We posit that whether that package of information is evaluated by itself—in separate evaluation (SE)—or side by side with another package—in joint evaluation (JE)—matters. Across a variety of contexts, four studies show that people "average" in SE and "add" in JE. Consequently, mildly favorable information *hurts* evaluations in SE but *helps* in JE. Study 1 demonstrated this "adding-and-averaging effect" among persons with expertise: law professors judging law faculty candidates. Adding middle tier academic publications to a higher tier publication on a CV *decreased* evaluations of a candidate judged in SE but *increased* evaluations of the same candidate in JE. Study 3 examined a linear pattern prediction, showing that each piece of mildly favorable information linearly *added* to the overall impression of a package of information in JE but linearly *detracted* from evaluations of the identical target in SE. Finally, Study 4 traced these differences in evaluative judgments to a shift in reference points brought about by evaluation mode. Implications for the organization specifically and our understanding of judgment and decision making processes more generally are considered.

Keywords: separate and joint evaluation, product bundling, evaluation mode, presenter's paradox, marketing

Managing impressions is relevant at multiple levels of consumer and organizational decision making, from incentivizing employees, to evaluating job applicants, from the marketing and distribution of products and services, to the organization at large, and this task has consequences (Podolny, 2010; Turnley & Bolino, 2001; Wayne & Liden, 1995). After all, impressions can determine whether an individual is hired into an organization, whether an organization is successful in a particular public relations campaign, and whether one's products and services garner market share. And one of the most robust findings in the literature on impression formation is that of weighted averaging in judgment. Even though adding mildly favorable information to highly favorable increases the total amount of positive information available about a stimulus, it has repeatedly been shown to lower judgments via the weighted averaging mechanism. This "averaging effect" has been shown to occur in a variety of contexts, ranging from the evaluation of product bundles and public relations campaigns (Weaver, Garcia, & Schwarz, 2012; Weaver, Hock, & Garcia, 2016) to the evaluation of performance incentives (Jagacinski, 1995) and products and services (Troutman & Shanteau, 1976; Weaver et al., 2012). In one prototypical demonstration, for instance, researchers showed that offering a small incentive, such as a $15 certificate for textbooks, alongside a large incentive such as a $1,750 tuition scholarship, actually *reduced* people's motivation to exert effort to obtain a similar reward in the future compared with a case where the identical higher-value incentive was offered alone. Like combining warm water with hot water produces water of an intermediate temperature, mildly favorable or lower-value items dilute the impact of highly favorable or higher cost ones, leaving perceivers with an overall impression that falls somewhere in the middle.

Given the importance of information presentation in both professional and everyday decision-making contexts (e.g., Highhouse, 1996; Jagacinski, 1991; Jenkins, Mitra, Gupta, & Shaw, 1998; Nagy, 1982; Slaughter, Bagger, & Li, 2006), understanding the process by which people form impressions of targets with multiple features is important to our understanding of judgment and decision making generally. In the current paper, we examine a potential moderator of the averaging effect that is of particular importance in organizational and consumer contexts—the evaluation mode in which a target is evaluated (Hsee, 1996, 1998; Hsee & Zhang, 2010; see also Bazerman, Loewenstein, & White, 1992). People typically evaluate options in one of two evaluation modes—by evaluating a target by itself in SE, as when a consumer decides whether to purchase one particular stereo system or a human resources representative decides whether or not to hire one particular applicant—or by evaluating a target in comparison with another target in JE mode, as when a consumer compares several

Kimberlee Weaver, Department of Business Administration, University of Haifa; Stephen M. Garcia, Department of Psychology, University of Michigan.

Correspondence concerning this article should be addressed to Kimberlee Weaver, Department of Business Administration, Faculty of Management, University of Haifa, Mount Carmel, Haifa 31905, Israel. E-mail: kweaver@univ.haifa.ac.il

potential stereo systems to each other or considers the qualifications of multiple applicants before choosing which one to hire. In the current paper, we hypothesize that averaging-like effects are more likely to occur in SE contexts, and that they give way to an "adding effect" under JE.

Averaging in Impression Formation

The original studies that examined whether people use adding-like or averaging-like processes when forming impressions of targets with multiple attributes were conducted in cognitive psychology in the context of person perception. This research showed that perceivers consistently gave more negative ratings to stimulus persons described by a combination of highly favorable and mildly favorable trait descriptors (H^+ H^+ H^+ M^+ M^+ M^+) than they gave to targets who had been described by the smaller number of highly favorable traits alone (H^+ H^+ H^+; e.g., Anderson, 1965; see Eagly & Chaiken, 1993 for a review). That is, even though adding three mildly favorable trait descriptors to a profile that already contained three highly favorable ones increased the total sum of positive information that perceivers had about a target, it lowered their evaluations of that person relative to targets described by the three highly favorable traits alone.

More recent studies have demonstrated that averaging-like processes are pervasive in consumer and organizational contexts as well. For example, when evaluating resumes where test scores of management skills and computer programming were positively correlated, research participants tended to average the scores together when making assessments of resumes (Jagacinski, 1995). Similarly, research participants rated the favorability of an academic job candidate with only one top tier publication as more impressive than did participants who rated a resume with that same top tier publication plus two middle tier publications (Weaver & Garcia, 2005). People also exhibit averaging-like processes when evaluating a variety of targets in consumer and every day decision making settings including arguments in public relations campaigns (Weaver et al., 2016), consumer products (Troutman & Shanteau, 1976; Weaver et al., 2012), and even evaluations of opposite sex coffee date partners based on pictures and personality information (Lampel & Anderson, 1968). In sum, the literature is replete with consistent and robust findings demonstrating averaging-like processes in human judgment across a variety of important contexts.

When Do "Add-Ons" Help in Evaluative Judgments?

But, when does mildly favorable information actually *add* value rather than *detract* from an overall evaluation? Interestingly, past work on averaging has been conducted almost exclusively in SE—contexts where people evaluate one target by itself. In the present analysis, we posit that the evaluation mode in which people make their evaluations—SE versus JE (Hsee, Loewenstein, Blount, & Bazerman, 1999; Hsee & Zhang, 2010)—is an important determinant of whether decision makers will average versus add when making evaluative judgments. SE refers to situations where people evaluate choice alternatives one at a time, whereas JE refers to comparative judgment contexts where people consider multiple alternatives simultaneously.

Rather than focusing on adding and averaging, past work on joint and separate evaluation (Hsee, 1996; Hsee et al., 1999; Hsee & Zhang, 2010) has emphasized *evaluability*: the idea that things that are noticed and differentiated in JE are often not noticed and differentiated in SE. Whether a dinnerware set has 24 or 33 intact pieces, for instance, is differentiated and noticed when the two sets are side by side in JE. In contrast, when people see only one of the sets in SE, they are not really sure how good 24 or 33 pieces actually is without a comparison standard—the number of pieces is "inevaluable"—and they are willing to pay (WTP) approximately the same for both. Extrapolating this finding to the consumer or resume context described above suggests that mildly favorable information in a product description or on a resume, for example, may add in JE but may produce a null effect in SE. This is distinct from our current hypothesis, which predicts that the *same information* that adds in JE can end up *detracting* in SE.

Where work on evaluability has shown actual *reversals* of preference between JE and SE, it is generally not due to the *same* information adding in JE and detracting in SE, but is instead due to extraneous patently negative features that are present in the experimental design (e.g., a torn cover of a book, broken dishware, or an offensive sound). These negative features overwhelm the theoretically predicted "null" or attenuated effect of the unevaluable item described above to produce a reversal of preference. For example, Hsee (1998) showed participants two sets of dinnerware like those described above. In JE, participants were WTP more for a set with 33 intact pieces and seven broken ones than for a set with 24 intact pieces only. However, when people saw only one of the sets in SE, they were actually WTP more for the smaller, perfect set than for the larger set with broken pieces. This reversal occurred not because the same thing that added in JE (the 9 extra intact pieces included in Set 1) ended up detracting in SE, but instead because the overtly negative information that one set contained broken pieces overwhelmed the theoretically predicted "null effect" in SE to create a preference reversal. However, a consistently and obviously negative piece of information is fundamentally different from a mildly positive one and does not have the same kind of consequences for self-presentational choices. For example, it is unlikely that an applicant would include clearly negative information—a broken glass so to speak—on a resume.

Thus, to summarize, our theoretical contribution is distinct in several important ways from past work on evaluability (Hsee et al., 1999; Hsee & Zhang, 2010). That work has shown that clearly negative attributes (e.g., a torn cover of a dictionary, broken dinnerware) can overwhelm unevaluable attributes (e.g., the number of entries in the dictionary, the total number of dishware pieces) in SE, but then receive less emphasis in JE, where previously uninterprerable dimensions—for example, the number of words in a dictionary—now take over in emphasis. In contrast, our adding-and-averaging hypothesis predicts a preference reversal of a different kind—that the *identical mildly favorable* feature that *detracts* from evaluations in SE mode will actually be the very thing that ends up adding value to evaluations in JE mode (cf., the torn cover or broken cups never added in previous work). Moreover, and importantly, the previous work on evaluability, unlike the present analysis, has made no claims about separate versus joint evaluation differentially prompting "adding" versus "averaging." So, our current question complements this past work by examining whether the same thing that adds in JE—mildly favorable information of the type that someone would include on a resume—can ever detract in SE.

That said, we hasten to acknowledge one pertinent study from economics that is not widely known in the organizational behavior literature. List (2002) presented people with either one set (SE) or two sets (JE) of baseball cards and gauged their willingness to pay (WTP) for the set. Set A contained 10 Topps baseball cards professionally graded to be in mint condition. Set B contained the same 10 cards plus three cards professionally graded to be "poor." Results showed that participants were WTP more for the 13-card set in JE but were WTP more for the 10-card set in SE. While relevant to the present analysis, there are important differences between that study and the present work.

First, rather than articulating an adding and averaging hypothesis to describe his results, List (2002) frames his study as being structurally similar to Hsee's (1998) dinnerware study. He even incorrectly states that Hsee's dinnerware sets contained "24 high quality pieces" and "24 high quality pieces and 16 low quality pieces." However, the second dinnerware set actually had 31 intact pieces and nine broken ones (i.e., not just "low quality" but arguably worthless to most people). Second, List's low quality baseball cards were labeled poor—which was the worst possible card grade. Thus, it is unclear whether that label had a heuristic negative effect on people's evaluations in SE. For instance, cards receiving the poor designation can be torn, stained, exhibit "noticeable warping," and include other types of extreme damage such as "extreme discoloration or dirtiness throughout that may make it difficult to identify the . . . content of the card" (Professional Sports Authenticator [PSA], 2018), which may provide salient heuristic negative information in evaluations. The fact that the mildly favorable information in his study is always paired with a damaged physical appearance leaves it ambiguous from a psychological standpoint why the preference reversal is occurring--is it due to the mildly favorable information per se or instead to a heuristic negative cue brought about by the extreme damaged appearance? This ambiguity did not interfere with List's (2002) goal of replicating Hsee's (1998) results in a market setting. However, a convincing test of the adding and averaging hypothesis advanced here requires a clean manipulation of mildly favorable information. After all, when trying to create a positive impression, there is no dilemma on whether unfavorable information should be added.

The Current Research

Building an adding and averaging account of organizational decisions, the present analysis complements previous work and thus posits the following hypotheses:

Hypothesis 1: Mildly favorable information will detract from evaluations when a bundle of information like a resume or product bundle is evaluated by itself in SE mode, but the identical mildly favorable information will produce an adding-like pattern when the same target is evaluated in the context of other candidates or products in JE mode.

Hypothesis 2: Averaging and adding-like patterns of evaluation occur because the JE and SE contexts produce shifting reference points, with SE prompting evaluators to compare to a reference *within* a given target, initiating an anchoring and adjustment process that leads to an averaging pattern, and JE prompting evaluators to compare to a reference *between* targets, producing between-target rather than within-target comparisons.

We tested our hypotheses in four studies. Our main prediction was that mildly favorable information would lower people's evaluations when targets were judged separately, but that the identical mildly favorable information would increase people's evaluations of the same targets when they were judged alongside other targets in JE. Study 1 reveals the reversal of preference among experts (i.e., law faculty judging publication records of Law faculty candidates). Study 2 replicates the effect among people with hiring experience. Study 3 shows that the effect also transpires in a multioption setting, producing a linear pattern of decreasing evaluations in SE and a linear pattern of increasing evaluations in JE. Study 4 offers evidence of a shifting reference point mechanism.

Study 1: The Adding-and-Averaging Effect in Evaluations: Expert Judgments

Study 1 examined our predictions in the context of a candidate for a law school faculty position. In order to examine whether expertise would moderate the effect, Law School faculty members at New York University, the University of California Los Angeles (UCLA), Northwestern University, and the University of Michigan (experts) and Amazon Mturk participants (nonexperts) judged the impressiveness of the academic publication record of a candidate (2 candidates) for a law school faculty position. One candidate had one top tier publication in the Columbia Law Review (ranked #4) and the other candidate also had a Columbia Law Review article as well as two additional mildly favorable articles (Kansas Law Review ranked #51 and Nebraska Law Review ranked #52). We predicted that including the publications from the middle tier journals would *hurt* the evaluations of the candidate in SE but would *help* those same evaluations when the candidate was evaluated in JE.

Will expertise matter? Past work on evaluability suggests that experts may be less likely to exhibit decision making biases than novices. Hsee and Zhang (2010), for instance, argued that experts should be *less* susceptible to joint/separate effects and that evaluation mode will be most likely to affect judgments when knowledge of the topic area is low. Work in the applicant selection literature has also shown that experts sometimes show fewer decision-making biases than novices. Expert recruiters, for instance, are less affected by job candidates' physical attractiveness (Marlowe, Schneider, & Nelson, 1996; Nagy, 1982; Shanteau, Thomas, Friel, Weiss, & Pounds, 2001), expert auditors and accountants are less susceptible to framing effects (Jamal, Johnson, & Berryman, 1995; Johnson, Jamal, & Berryman, 1991), and expert managers are less likely to show the halo error than those with less experience (Kozlowski & Kirsch, 1987).

One reason why experts may show fewer decision making biases is the result of training. Expert managers, for instance, may have been trained to ignore physical characteristics in candidate evaluations, and formal training also reduces biases like the halo error (Athey & McIntyre, 1987; Borman, 1975). The adding- and averaging-like processes hypothesized here, in contrast, are less well-known, and the anchoring and adjustment process, which we posit underlies averaging-like patterns, is a basic human perception process and thus may be less correctable. Indeed, past work in the

applied psychology literature shows that biases arising from basic processes tend to be less amenable to training (e.g., increased thought does not eradicate the contrast effect: Wexley, Sanders, & Yukel, 1973; see also LeBoeuf & Shafir, 2003, increased thought does not reduce framing effects; Tversky & Kahneman, 1974, experts are susceptible to representativeness). For these reasons, we predicted that experts would be as susceptible to the adding-and-averaging effect when evaluating the candidates' publication records as novices.

Participants and Procedure

We emailed the law faculty of NYU, Northwestern University, UCLA, and the University of Michigan: 104 law faculty responded (33% female, 19.3% response rate). For our nonexpert sample, 101 (44% female) participants were recruited from Amazon MTurk. Expertise was thus a between-subjects variable. All participants read:

> Imagine you are serving on a search committee to recruit a junior Law School faculty member. Candidates publish in journals called Law Reviews and each is ranked from 1-135, with 1 'the best' and 135 'the worst.' Here is the publication[s] submitted by Candidate A.

Participants in the SE condition then saw one candidate in a between-subjects design who had either published one article, "Columbia Law Review (#4 rank)" or three articles "Columbia Law Review (#4 rank)," "Kansas Law Review (#51)" and "Nebraska Law Review (#52)," and then responded to the dependent variable, "Based on this information, how impressive is Candidate A?" (1 = *unimpressive*, 7 = *impressive*). Participants in JE saw both profiles side by side in a within-subjects design as "Candidates A and B" (order counterbalanced), and responded to the identical "impressiveness" dependent variable described above for both candidates. Separate groups of participants were run to validate the stimulus materials for this and all studies, see the Appendix.

Results and Discussion

There was no interaction between university and rated impressiveness for either the JE or SE conditions (all $ps > .31$), so this variable is not discussed further. To analyze the results, two separate 2 × 2 analyses of variance (ANOVAs) were run, one for the SE conditions and another for the JE conditions: a 2 (expertise: expert vs. nonexpert) × 2 (number of articles published: Columbia only vs. Columbia plus Kansas and Nebraska) between-subjects ANOVA for SE and a 2 (expertise: expert vs. nonexpert) × 2 (number of articles published: Columbia only vs. Columbia plus Kansas and Nebraska) mixed-model ANOVA (with number of articles published now a within-subjects factor) for the JE condition.

As predicted, results showed that when evaluators judged only one of the candidates at a time in SE, their evaluations showed an averaging-like pattern: they were significantly more impressed by the candidate with a single Columbia Law Review article ($M = 5.71$, $SD = 1.16$) than by the candidate with a publication in that same journal plus two additional articles in middle-tier journals ($M = 4.84$; $SD = 1.14$), $F(1, 99) = 15.98$, $d = .76$, $p < .001$. This pattern was identical and individually significant regardless of whether the evaluators were experts: faculty members (Columbia only: ($M = 5.16$, $SD = 1.11$; Columbia plus: $M = 4.60$, $SD = 1.16$), $F(1, 42)$, $d = .49$, $p < .05$ or nonexperts (Columbia only: $M = 6.22$; $SD = .97$; Columbia plus: $M = 5.08$, $SD = 1.09$), $F(1, 51) = 16.3$, $d = 1.10$, $p < .001$).

However, as predicted, in JE the pattern was the opposite: now evaluators judged the candidate with the *three* publications to be significantly more impressive ($M = 5.17$; $SD = 1.13$) than the candidate with the Columbia Law Review article only ($M = 4.49$; $SD = 1.20$, $F(1, 96) = 20.78$, $d = .58$, $p < .0001$). This pattern was again identical and individually significant regardless of whether the evaluators were faculty members (Columbia only: $M = 4.26$, $SD = 1.18$, Columbia plus: $M = 5.04$, $SD = 1.14$) within-subjects ANOVA, $F(1, 49) = 20.0$, $d = .67$, $p < .001$ or nonexperts (Columbia only: $M = 4.73$, $SD = 1.18$; Columbia plus: $M = 5.31$, $SD = 1.11$), within-subjects ANOVA, $F(1, 47) = 5.66$, $d = .51$, $p < .05$). There were no interactions with expertise in either of the evaluation modes ($ps > .17$). Figure 1A displays the pattern of results for the expert sample and Figure 1B for the nonexpert sample.

These results demonstrate for the first time that experts show a significant adding-and-averaging effect across JE and SE: law

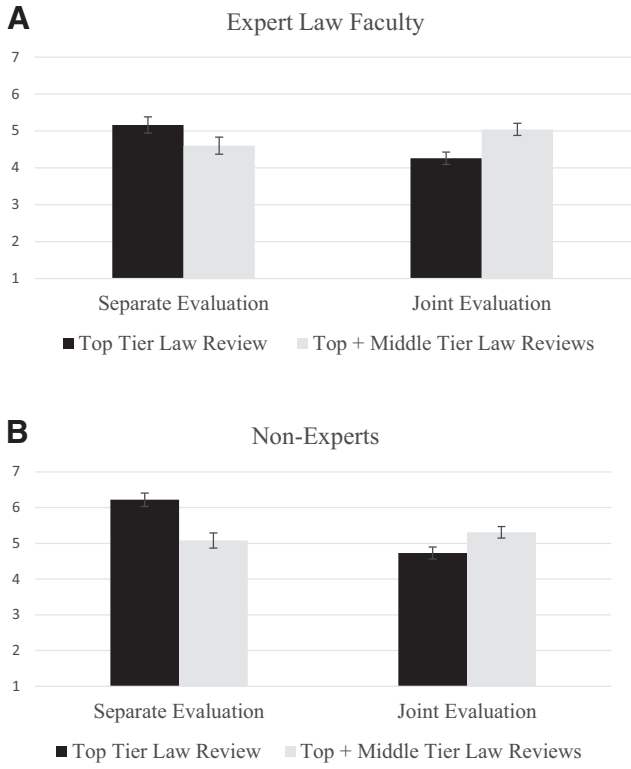

Figure 1. (A) Expert law faculty members' perceived impressiveness of the candidates (1 = *not at all impressive*, 7 = *very impressive*) in Study 1 (shown with standard error bars) as a function of evaluation mode and whether the candidate had included the two middle tier publications in the CV. (B) Nonexperts' perceived impressiveness of the candidates (1 = *not at all impressive*, 7 = *very impressive*) in Study 1 (shown with standard error bars) as a function of evaluation mode and whether the candidate had included the two middle tier publications in the CV.

faculty members in SE rated a candidate with a single top tier publication as more impressive than a candidate with a similar top tier publication plus two additional middle tier publications. However, law faculty members in JE showed the opposite pattern of evaluations: now the candidate with one top tier and two middle tier publications was rated as significantly *more* impressive than the candidate with one top tier publication only.

Study 2: Adding Versus Averaging and Previous Hiring Experience

Study 2 conceptually replicated the findings of Study 1 using a different resume context with the mildly favorable information being a second major with a less than perfect GPA. We predicted that the second major would *hurt* the candidate in SE, but would *help* him or her in JE, even among evaluators with past hiring experience.

Participants and Procedure

Participants ($N = 290$; 34% female, $M_{age} = 30.6$; range 18–61) were recruited from Amazon MTurk and the alumni of a large state university and read, "Imagine that you are hiring for an entry-level position. You are searching for recent college grads. Here is the educational background . . ." Participants in the SE condition then saw information about Candidate A in a two-condition between-subjects design either as "Major: Business (GPA: 3.98)" or "Double Major: Business (GPA: 3.98) and Computer Science (GPA 2.95)" and reported, "Based on this information, how impressive is Candidate A?" (1 = *unimpressive*, 7 = *impressive*). Participants in JE saw both candidates side by side in a within-subjects design as Candidates A and B (counterbalanced) and rated the impressiveness of each. Participants were then asked, "Have you ever served on a hiring committee or made a hiring decision?" reported their industry, and indicated whether or not they take educational background into account when hiring recent college graduates.

Results and Discussion

Participants represented a wide range of industries (e.g., insurance, health care, nonprofits, engineering) and 48% reported having previous hiring experience, of which 92% reported taking into account educational background when hiring recent college graduates. Type of sample (MTurk vs. university alumni) did not interact with experimental condition ($ps > .2$), so the samples were combined. Note that because of the experimental designs we use a fully between-subjects ANOVA to analyze the data in SE and a mixed-model ANOVA (with presence vs. absence of the double major now a within-subjects factor) to analyze the data in JE.

As predicted, regardless of hiring experience, participants in SE perceived the candidate with the 3.98 GPA business major only to be more impressive ($M = 5.70$, $SD = 1.13$) than the candidate with both the 3.98 GPA business major plus the additional 2.95 GPA computer science major ($M = 5.00$, $SD = 1.13$, between-subjects ANOVA $F(1, 145) = 14.19$, $d = .62$, $p < .001$). However, in JE, we observe a reversal: the candidate with the 3.98 GPA business major plus the 2.95 GPA computer science major now is perceived to be more impressive ($M = 5.71$, $SD = 1.14$) than the candidate with the 3.98 GPA business major alone ($M = 5.06$, $SD = 1.10$, within-subjects ANOVA $F(1, 143) = 34.19$, $d = .58$, $p < .001$).

The same pattern is obtained when we examine the results separately by hiring experience. Participants with actual hiring experience were more impressed by the candidate with the business major alone ($M = 5.61$, $SD = 1.10$) than by the candidate who had the identical business degree along with an additional double major in computer science with the less impressive GPA ($M = 4.97$, $SD = 1.19$), $F(1, 67) = 5.42$, $d = .56$, $p < .05$. However, as predicted, in JE, participants with actual hiring experience showed the opposite pattern: the candidate with the double major was now seen to be significantly more impressive ($M = 5.62$, $SD = 1.05$) than the candidate with the business major alone ($M = 5.14$, $SD = 1.00$), $F(1, 70) = 8.04$, $d = .47$, $p < .01$. The same pattern of results emerges for those without hiring experience: SE (business major alone: $M = 5.79$, $SD = 1.17$ versus double major: $M = 5.03$, $SD = 1.10$, $F(1, 76) = 8.90$, $d = .67$, $p < .01$) and JE (double major: $M = 5.79$, $SD = 1.22$ versus business major alone $M = 4.97$, $SD = 1.19$, $F(1, 72) = 32.1$, $d = .68$, $p < .001$). These results thus conceptually replicate among those with previous hiring experience the adding-and-averaging effect pattern shown in Study 1. Figure 2A displays the pattern of results for the sample with hiring experience and Figure 2B for the sample without hiring experience.

Study 3: Multi-Option Study

Another phase of the hiring experience is the recruitment phase. Using three options, we predicted that each mildly favorable "add-on" to a salary package would increasingly *diminish* the impressiveness of the package overall in SE yet would increasingly *boost* the impressiveness of the same package in JE. We note that this adding-averaging linear pattern has not been previously shown in a multi-option study. In fact, in one previous study (Hsee & Zhang, 2004), participants' happiness ratings linearly increased with the number of times their books sold in JE (i.e., 0, 80, 160, or 240 copies). However, in SE, participants were "value insensitive"—the linear relationship between happiness and number of books sold was attenuated but still in the same direction. In the present study, our prediction is that the mildly favorable information that linearly adds in JE will linearly *detract* in SE.

Participants and Procedure

One hundred and one participants based in the United States (40.6% female; $M_{age} = 30.36$; range 18–61) were recruited from Amazon MTurk and read, "Imagine that you are about to graduate from college, and have been offered the following salary package from Company A in the financial district of San Francisco." Participants in SE then saw either "$58,000 salary," that same salary and a "$10 monthly voucher toward gym membership," or the same salary, the gym voucher, and a "$10 monthly voucher toward parking expenses" (between-subjects). The dependent variable was "How impressive is the salary package of Company A" (1 = *unimpressive*, 7 = *impressive*). Participants in JE rated the impressiveness of all three salary packages from "Companies A, B, and C" in one of two counterbalanced orders (i.e., Version 1: salary alone, salary + gym, salary + gym + parking; Version 2: the reverse; within subjects).

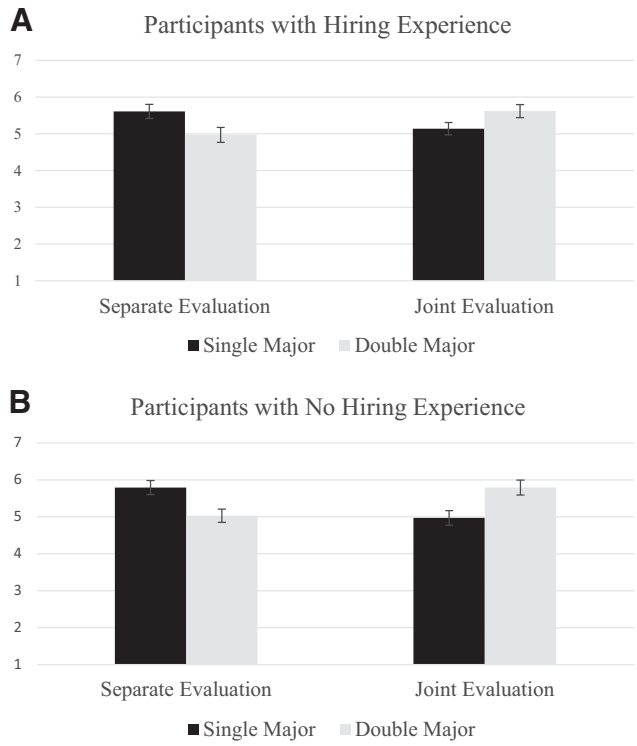

Figure 2. (A) Among participants with hiring experience, the perceived impressiveness of the candidate (1 = *unimpressive*, 7 = *impressive*) in Study 2 (shown with standard error bars) as a function of the evaluation mode and whether the candidate has a single major in business (3.98 GPA) or a double major in business (3.98 GPA) and computer science (2.95 GPA). (B) Among participants with no hiring experience, the perceived impressiveness of the candidate (1 = *unimpressive*, 7 = *impressive*) in Study 2 (shown with standard error bars) as a function of the evaluation mode and whether the candidate has a single major in business (3.98 GPA) or a double major in business (3.98 GPA) and computer science (2.95 GPA).

Results and Discussion

Results supported predictions. Participants in SE exhibited a pattern that resembled averaging; they were increasingly less impressed with each additional mildly favorable add-on to the package: salary alone ($M = 4.65$, $SD = 1.15$), salary + gym membership ($M = 4.00$, $SD = 1.45$), and salary + gym + parking ($M = 3.55$, $SD = 1.91$). A between-subjects linear contrast (+1, 0, −1) confirmed this pattern was significant ($F(1, 58) = 6.03$, $p < .05$); the orthogonal contrast (−1 +2 −1) was not, $F(1, 58) = .07$, $p > .79$. In JE, the exact opposite pattern emerged instead; participants now became increasingly more impressed with each mildly favorable add-on to the package: salary only ($M = 4.21$, $SD = 1.42$), salary + gym ($M = 4.79$, $SD = 1.15$), and salary + gym + parking ($M = 5.23$, $SD = 1.32$). A within-subjects linear contrast (−1, 0, +1) showed this pattern was significant, $F(1, 84) = 19.2$, $p < .001$; the orthogonal contrast (−1 +2 −1) was not, $F(1, 84) = .12$, $p > .72$. Study 3 thus shows that the averaging and adding-like patterns extend beyond the two option targets investigated in past work to multi-option ones.

Study 4: Shifting Reference Points

We predicted that the two evaluation modes produce different patterns of judgment because they make different reference points salient. Study 4 tested this hypothesis in the context of hiring a consultant firm. We hypothesized that in SE evaluators would use attributes of the firm itself—that is, the ranks of its past clients—as reference points for each other, leading the higher ranked clients to make the lower ranked look worse, producing an anchoring and adjustment process that leads to an averaging pattern of judgment. In contrast, we predicted that the reference point in JE would shift from the individual attributes within a given candidate to differences between the candidates. Thus, in JE the lower ranked previous clients will be compared with the other firm's absence of additional clients. While Hsee and Zhang (2010) also use a reference point account to describe evaluability effects in JE and SE, reference points in SE were typically internal (i.e., knowledge), external (i.e., standard), or neutral. Here we suggest that attributes of the target itself within the SE context can also reflect reference points.

Participants and Procedure

Participants ($N = 108$; 40% female) were recruited from MTurk and read:

> Imagine that you are the chief operating officer of a bank, and you are looking to hire a consulting firm to help streamline your operations. One consulting firm—Candidate Firm 1—has provided an example of their past clients in the banking industry.

Participants in SE then saw in a between-subjects design either "Wells Fargo Bank (#1 ranked)"—or "Wells Fargo Bank (#1 ranked), Hudson City Bancorp (#47 ranked), and First Citizens BancShares (#67 ranked)," and were asked, "How good is Candidate Firm 1?" (1 = *not very*, 7 = *very*)." Participants in JE saw both firms side by side in a within-subjects design and rated both of them on the identical dependent measure. To capture the postulated shifting reference point process, participants then reported how they made their evaluations in an open-ended fashion.[1] Two judges blind to hypotheses coded the responses (85.2% agreement) as to whether they focused on the rank of the top tier client as a reference point and noted comparisons of the other client to that top tier client (e.g., "They are pretty good since they worked with a #1 ranked bank but the other two are very much lower"; "They have one really good client and 2 mediocre clients so I rated them higher than average but not the best"; "#1 bank was a client but the others are much lower rated") or on the other firm as a reference point, without reference to the individual ranks of the clients and how they compared with one another (e.g., "Firm 2 is higher because they had more clients"; "Firm 2 clearly has more experience, because it has had more clients"; "Firm 2 has more of a client list than firm 1, and has the same high name client").

[1] Although Nisbett and Wilson (1977) have highlighted issues with free response data, their review paper largely focused on difficult to assess phenomena related to cognitive dissonance, arousal, and subliminal priming. Free response data are of value (Haddock & Zanna, 1998), and, in the present analysis, participants should more readily have access to the reference point account that we seek to probe.

Participants evaluating the three-client firm (the SE/3 client and the JE conditions) also reported whether they took into account the number of past clients as well as how their judgments would have changed if more second tier clients had been added to the firm's list "How much more favorably would you have rated the [three client] firm if they had had two additional clients ranked #68 and #70?" (1 = *much less favorably*; 7 = *much more favorably*). If the reference point used was the other firm, then adding two additional mildly favorable clients should increase evaluations, adding to their judgments. In contrast, if the reference point participants used to make their evaluations was the top tier client itself, then adding two additional mildly favorable clients should not increase their evaluations in a positive way.

Results and Discussion

As predicted, a between-subjects design ANOVA showed that participants in SE exhibited an averaging pattern, rating the consulting firm advertising one highly ranked client as better (M = 5.53, SD = 1.30) than the firm noting that same highly ranked client along with two lower ranked ones (M = 4.86, SD = 1.13, $F(1, 64) = 5.0$, $d = .40$, $p < .05$). As predicted, participants in the JE condition, on the contrary, exhibited the opposite pattern: now the firm who advertised three example clients was evaluated as significantly better (M = 5.28, SD = 1.10) than the firm who submitted only one (M = 4.49, SD = 1.30, $F(1, 38) = 9.15$, d = .66, $p < .005$) (within-subjects).

The additional measures supported the reference point account: more participants in *SE* referenced the ranking of the top tier firm as a reference point, noting that the two lower ranked firms were less prestigious, and stating that they adjusted their evaluations of the firm accordingly (64.8%) than JE participants (25.8%), χ^2 = 10.32, $p < .01$. In contrast, more participants in JE mentioned the other firm as a reference point and focused on the larger number of clients as the basis for their evaluation without referencing the individual ranks of the firms themselves (JE: 86.4%; SE: 26.1%), $\chi^2 = 21.8$, $p < .001$. The quantitative measures mirrored these results. Adding two clients ranked #68 and #70 increased JE participants' evaluation of the three-client firm (M = 4.58; SD = 2.22) but not SE participants' evaluations of the same portfolio (M = 3.21; SD = 1.58), $F(1, 62) = 8.17$, $d = .71$, $p < .006$. JE participants also reported attending to the number of clients more (M = 5.31; SD = 1.86) than SE participants rating the three-client firm (M = 2.71; SD = 1.68), $F(1, 64) = 35.85$, $d = 1.47$, $p <$ .0001.

General Discussion

Whether it is attempting to recruit the very best personnel for the organization, choosing what portfolio to submit in a call for consulting firms, designing a public relations campaign or product bundle to increase market share, or designing incentives to motivate performance, understanding when mildly favorable information adds versus detracts from the overall impression of a presentation is key. Using the broad conceptual framework of separate versus joint evaluation (e.g., Bazerman et al., 1992; Hsee, 1996; Hsee et al., 1999; Hsee & Zhang, 2010), we demonstrate evidence of an adding-and-averaging effect: the evaluation mode in which people make their decisions is a significant moderator of whether decision makers will exhibit adding- or averaging-like patterns of evaluation when forming evaluations. Specifically, adding mildly favorable attributes (e.g., middle tier journal publications, lower ranked firms) to highly favorable ones (e.g., a top tier publication, a #1 ranked firm) leads to averaging-like patterns of judgment and thus detracts from evaluations when a package of information is evaluated by itself in a SE context. In contrast, the identical mildly favorable information will actually add positivity to people's judgments of the overall favorability of a package when the same decision alternatives are presented alongside other options in JE.

Four studies provided support for this hypothesis and illustrate the quandary that presenters of information may find themselves in when trying to present themselves and their organizations in the best possible light. Study 1 showed that law faculty members and novices alike showed the adding-and-averaging effect when evaluating the scholarly publication records of prospective faculty candidates—candidates with additional mildly favorable publications were seen to be *less* impressive when evaluated in SE, but as *more* impressive when evaluated in JE. Study 2 likewise showed that a double major in computer science with a mediocre GPA *hurt* a job applicant in SE, but *helped* the same candidate in JE and that the adding-and-averaging pattern emerged regardless of evaluator hiring experience. Study 3 revealed these dynamics in a multioption setting, showing linear *decreases* in impressiveness with each additional piece of mildly favorable information in SE, but linear *increases* with the exact same information in JE. Finally, Study 4 probed for the mechanism of the effect, finding support for our shifting reference point account.

Theoretical Contributions

The current research contributes to our theoretical understanding of impression formation processes as well as to the literature on joint versus SE more generally. Previous work attempting to distinguish between adding-like and averaging-like processes in human judgment has found overwhelming support for averaging-like processes. At the same time, this past literature has inadvertently focused almost exclusively on contexts where the targets of evaluation are judged separately. By linking two previously distinct lines of research together—the literature on adding and averaging along with that on separate versus joint evaluation—the current research brings to light an important contextual factor that moderates when mildly favorable information will be added to or averaged into decision makers' judgments.

The present analysis also builds upon and complements previous research that has examined the effects of SE versus JE on the weighting of attributes in decision contexts (Hsee, 1996; Hsee et al., 1999; Hsee & Zhang, 2010). While this latter work has shown that manifestly *negative* attributes (e.g., a torn cover of a dictionary; broken dishware) often exert a large effect in SE mode but that their effect is mitigated in JE contexts, the present analysis demonstrates situations where the *very feature that detracts* from the evaluation in SE mode—mildly favorable information—actually becomes the very thing that ends up adding value in JE (cf., the torn cover or broken dishware in previous work never "added").

We further show that this effect occurs in contexts where the heuristic negative value of information is not salient (e.g., List, 2002).

The current research also shows that, contrary to predictions from evaluability theory and some past work in applied psychology, which generally predicts that experts should be less susceptible to evaluation mode effects and other decision biases, experts did show significant adding and averaging effects in the current contexts. In so doing, this work contributes to the literature by identifying an additional class of phenomena whereby experts, at least under some conditions, are susceptible to judgmental biases.[2] Finally, the current research also fills in theoretical gaps within the evaluation mode literature by explicitly linking evaluation mode to adding and averaging-like patterns, and by showing these dynamics in the multi-option setting for the first time.

Limitations and Future Directions

We have proposed a shifting reference points account of the mechanism. Namely, evaluators in SE use the individual attributes within a given candidate as reference points that lead to an averaging-like pattern. In JE, however, the reference point shifts from attributes within a given candidate to differences in attributes between the candidates. Indeed, Study 4 provides results consistent with a shifting reference point account in both qualitative responses as well as Likert-scale measures. For instance, in the SE condition, perceivers were more likely to mention the rankings of the previous clients and how they related to each other than were perceivers in JE, while the latter were more likely to specify the overall number of clients in explaining their evaluations. Despite linking a shifting reference points explanation to the present effects, the shifting reference points account could itself be explained by at least two different processes.

One possibility is that the different evaluation modes actually prompt different integration processes, with SE directly setting up a context that promotes averaging and JE a context that promotes adding. For instance, the natural emphasis on one single target in SE may lead perceivers to form a unified impression of the information, which may naturally lead them to attempt to blend the target's individual features into one coherent holistic picture (Weaver et al., 2012). This process may thus produce an effect that resembles—both on an empirical and on an actual cognitive process level—averaging. In contrast, the natural emphasis in JE is on the comparison of targets, which may lead perceivers to focus on individual features of each target and differences between them. This process may lead to a more piecemeal integration strategy that resembles adding (Weaver et al., 2012).

Another possible explanation is that rather than leading to different integration processes per se, the different evaluation modes may affect perceivers' perceived judgments of the favorability of the individual features comprising each target. According to this analysis, an article in the Kansas Law Review (ranked #51) may look less favorable in SE when a person also has a Columbia Law Review article (ranked #4) than it does in JE when the reference point is now another candidate without such an article. In this case, the integration rules themselves may be the same across SE and JE, but the favorability values of the pieces of information that are being integrated within the two modes may be different. Indeed, the scale distortion theory of anchoring (Frederick & Mochon, 2012) may also be relevant here.

While future research is obviously needed for an even finer-grained explanation of the cognitive processes behind the reference points process, future research is also needed to explore the extent to which evaluation mode in general can help address other theoretical and applied questions. For example, perhaps evaluation mode could provide a more cognitive explanation as to why the metacontrast ratio—where perceived ingroup differences are minimized and outgroup differences are maximized—emerges in social identity research. It may well be the case that one kind of self-categorization (Leonardelli & Toh, 2015) requires a JE context whereas another kind can emerge in an SE context. Evaluation mode may also be helpful in addressing the construct validity problem of assessment centers in the evaluation of personnel (e.g., Arthur, Day, McNelly, & Edens, 2003). For example, JE may help explain results observed when personnel are assessed together while SE may explain results that are observed when personnel are assessed individually. And finally, it would be helpful to establish the basic effects observed here using experimental paradigms that involve real consequences or high-stakes decisions. Doing so would show the robustness of the effect as well as contribute to the base of empirical evidence.

Practical Implications

In addition to theoretical contributions, these results also provide useful practical insights for self presenters and decision makers in general who are interested in how to best present information, although we acknowledge that the real world is undoubtedly messier and less compartmentalized, so to speak, than the designs of our experimental paradigms. Nevertheless, at the organizational level of analysis, companies may spend thousands of dollars a year on adding "bonuses" to employee performance incentives (i.e., membership to a mediocre gym) or even bundling "perks" or "add-ons" with high dollar products (e.g., voucher for free coffee). Therefore, knowing when those mildly favorable items will facilitate the goal, and when they will actually hurt it is useful information. Such companies would be well-advised to take into account whether such incentives and products are generally evaluated in joint versus SE contexts when making such incentivizing decisions. At the individual level of analysis, the same prescription applies: individuals pitching new product ideas to management, or even their own credentials for a promotion or job opening, or the products and services of the organization as a whole, should be mindful of the evaluation mode in which their presentations will be evaluated.

By showing that mildly favorable information can exert opposite effects on judgments in joint versus separate evaluation, the current research also brings to light the potential for some interesting paradoxes to arise as a function of the overall decision making context. For example, it is important to note that both organizational and consumer decision processes often utilize SEs and JEs at different stages in the decision sequence. For instance, in the context of job applications, there may be an initial stage where a large number of candidates or possibilities are evaluated in JE. Afterward, a much smaller number of individuals may be "short listed" and be invited for individual interviews, and the uniqueness of those visits may foster

[2] We do not deny that some kinds of experts might be impregnable to these effects. Experts who rely on highly specialized standards (e.g., diamond experts) or those who have been given explicit, well-defined standards, for instance, are probably less likely to succumb to these effects.

something more akin to SE mode. Interestingly, what may have benefited a job candidate in one phase may end up being detrimental in another. For instance, while including lower ranked publications on one's CV may help a given candidate make the short list (JE), once he or she makes it to the final group those same accomplishments may now detract (SE).

Conclusion

Building upon the evaluation mode literature, the present analysis posits that people average in SE and add in JE such that the identical mildly favorable information that hurts impressions in SE will actually help in JE. Given the ubiquity of evaluation mode, this adding-and-averaging effect inherently offers broad implications for individuals interested in how to most effectively present information to others in both professional and everyday decision-making contexts.

References

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology, 70,* 394–400. http://dx.doi.org/10.1037/h0022280

Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56,* 125–153. http://dx.doi.org/10.1111/j.1744-6570.2003.tb00146.x

Athey, T. R., & McIntyre, R. M. (1987). Effect of rater training on rater accuracy: Levels-of-processing theory and social facilitation theory perspectives. *Journal of Applied Psychology, 72,* 567–572. http://dx.doi.org/10.1037/0021-9010.72.4.567

Bazerman, M. H., Loewenstein, G. F., & White, S. B. (1992). Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives. *Administrative Science Quarterly, 37,* 220–240. http://dx.doi.org/10.2307/2393222

Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology, 60,* 556–560. http://dx.doi.org/10.1037/0021-9010.60.5.556

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General, 141,* 124–133. http://dx.doi.org/10.1037/a0024006

Haddock, G., & Zanna, M. P. (1998). On the use of open-ended measures to assess attitudinal components. *British Journal of Social Psychology, 37,* 129–149.

Highhouse, S. (1996). Context-dependent selection: Effects of decoy and phantom job candidates. *Organizational Behavior and Human Decision Processes, 65,* 68–76. http://dx.doi.org/10.1006/obhd.1996.0006

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67,* 247–257. http://dx.doi.org/10.1006/obhd.1996.0077

Hsee, C. K. (1998). Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making, 11,* 107–121. http://dx.doi.org/10.1002/(SICI)1099-0771(199806)11:2<107::AID-BDM292>3.0.CO;2-Y

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical Analysis. *Psychological Bulletin, 125,* 576–590. http://dx.doi.org/10.1037/0033-2909.125.5.576

Hsee, C. K., & Zhang, J. (2004). Distinction bias: Misprediction and mischoice due to joint evaluation. *Journal of Personality and Social Psychology, 86,* 680–695. http://dx.doi.org/10.1037/0022-3514.86.5.680

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

Jagacinski, C. M. (1991). Personnel decision making: The impact of missing information. *Journal of Applied Psychology, 76,* 19–30. http://dx.doi.org/10.1037/0021-9010.76.1.19

Jagacinski, C. M. (1995). Distinguishing adding and averaging models in a personnel selection task: When missing information matters. *Organizational Behavior and Human Decision Processes, 61,* 1–15. http://dx.doi.org/10.1006/obhd.1995.1001

Jamal, K., Johnson, P. E., & Berryman, R. (1995). Detecting framing effects in financial statements. *Contemporary Accounting Research, 12,* 85–105. http://dx.doi.org/10.1111/j.1911-3846.1995.tb00482.x

Jenkins, G. D., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of Applied Psychology, 83,* 777–787.

Johnson, P. E., Jamal, K., & Berryman, R. G. (1991). Effects of framing on auditor decisions. *Organizational Behavior and Human Decision Processes, 50,* 75–105. http://dx.doi.org/10.1016/0749-5978(91)90035-R

Kozlowski, S. W., & Kirsch, M. P. (1987). The systematic distortion hypothesis, halo, and accuracy: An individual-level analysis. *Journal of Applied Psychology, 72,* 252–261. http://dx.doi.org/10.1037/0021-9010.72.2.252

Lampel, A. K., & Anderson, N. H. (1968). Combining visual and verbal information in an impression-formation task. *Journal of Personality and Social Psychology, 9,* 1–6. http://dx.doi.org/10.1037/h0025716

LeBoeuf, R. A., & Shafir, E. (2003). Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making, 16,* 77–92. http://dx.doi.org/10.1002/bdm.433

Leonardelli, G. J., & Toh, S. M. (2015). Social categorization in intergroup contexts: Three kinds of self-categorization. *Social and Personality Psychology Compass, 9,* 69–87. http://dx.doi.org/10.1111/spc3.12150

List, J. A. (2002). Preference reversals of a different kind: The"more is less" phenomenon. *The American Economic Review, 92,* 1636–1643. http://dx.doi.org/10.1257/000282802762024692

Marlowe, C. M., Schneider, S. L., & Nelson, C. E. (1996). Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology, 81,* 11–21. http://dx.doi.org/10.1037/0021-9010.81.1.11

Nagy, G. F. (1982). *How are personnel selection decisions made? An analysis of decision strategies in a simulated personnel selection task* (Unpublished doctoral dissertation). Kansas State University, Manhattan, KS.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Podolny, J. M. (2010). *Status signals: A sociological study of market competition*. Princeton, NJ: University Press.

Professional Sports Authenticator (PSA). (2018, April 8). *Grading standards for baseball cards*. Retrieved from https://www.psacard.com/resources/gradingstandards/#detail

Shanteau, J., Thomas, R. P., Friel, B., Weiss, D. J., & Pounds, J. C. (2001). *Identifying expertise without a gold standard: Four applications*. Paper presented at the 11th International Symposium on Aviation Psychology, Columbus, OH.

Slaughter, J. E., Bagger, J., & Li, A. (2006). Context effects on group-based employee selection decisions. *Organizational Behavior and Human Decision Processes, 100,* 47–59. http://dx.doi.org/10.1016/j.obhdp.2006.01.003

Troutman, M. C., & Shanteau, J. (1976). Do consumers evaluate products by adding or averaging attribute information? *Journal of Consumer Research, 3,* 101–106. http://dx.doi.org/10.1086/208657

Turnley, W. H., & Bolino, M. C. (2001). Achieving desired images while avoiding undesired images: Exploring the role of self-monitoring in impression management. *Journal of Applied Psychology, 86,* 351–360. http://dx.doi.org/10.1037/0021-9010.86.2.351

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. http://dx.doi.org/10.1126/science.185.4157.1124

Wayne, S. J., & Liden, R. C. (1995). Effects of impression management on performance ratings: A longitudinal study. *The Academy of Management Journal, 38,* 232–260. http://dx.doi.org/10.2307/256734

Weaver, K., & Garcia, S. M. (2005). The self presenter's paradox: Motivated reasoning in impression formation. *Association for Consumer Research Geeta Menon and Ashkay Rao (Eds.), 32,* 363–364.

Weaver, K. D., Garcia, S. M., & Schwarz, N. (2012). The presenter's paradox. *Journal of Consumer Research, 39,* 445–460. http://dx.doi.org/10.1086/664497

Weaver, K., Hock, S. J., & Garcia, S. M. (2016). "Top 10" reasons: When adding persuasive arguments reduces persuasion. *Marketing Letters, 27,* 27–38. http://dx.doi.org/10.1007/s11002-014-9286-1

Wexley, K. N., Sanders, R. E., & Yukel, G. A. (1973). Training interviewers to eliminate contrast effects in employment interviews. *Journal of Applied Psychology, 57,* 233–236. http://dx.doi.org/10.1037/h0034714

Appendix

Stimulus Evaluations

Because we manipulate highly favorable versus mildly favorable information, we performed manipulation checks with separate groups of participants, in between-subjects design tests, to validate the favorability of the stimuli used in our studies. The manipulation checks indeed corroborate these naturally intuitive manipulations. For Study 1 ($N = 87$), the Columbia Law Review publication was perceived as being more impressive ($M = 6.52$; $SD = .87$) than the Kansas ($M = 4.68$; $SD = 1.25$) and Nebraska ($M = 4.4$; $SD = 1.33$) law review publications, $F(1, 84) = 35.31$, $p < .001$; $F(1, 84) = 38.13$, $p < .001$. The two lower ranked journals themselves were perceived as more impressive than the scale point that had been labeled as neutral on the scale (4 = *neither impressive nor unimpressive*; one-sample $t(27) = 4.99$, $p < .001$; $t(30) = 4.40$, $p = .001$, respectively). For Study 2 ($N = 57$), the 3.98 GPA business major was perceived as more impressive ($M = 6.25$; $SD = .80$) than the 2.95 GPA computer science major ($M = 4.41$; $SD = 1.06$), $F(1, 55) = 71.77$, $p < .001$. The 2.95 computer science major was judged to be mildly favorable in and of itself compared with the neutral point on the scale (4 = *neither impressive nor unimpressive*; one-sample $t(28) = 3.24$, $p = .003$). For Study 3 ($N = 87$), the $58,000 salary was perceived as more favorable than the $10 monthly gym voucher and $10 parking pass ($M = 2.61$; $SD = .84$ vs. $M = 2.23$; $SD = .83$), $F(1, 85) = 4.14$, $p < .05$, and the two mildly favorable perks were themselves rated as more favorable than the neutral point (one sample t test compared with "2" on the 4-point scale labeled "mildly favorable," $t(55) = 2.09$, $p = .04$). For Study 4 ($N = 78$), the #1-ranked Wells Fargo Bank was rated as more favorable than #47 Hudson City Bancorp and #67 First Citizen Bancshares banks, $F(1, 76) = 3.86$, $p = .05$, and the two mildly favorable banks were themselves seen as more favorable than the neutral point (one-sample t test compared with "2" on the 4-point scale labeled *mildly favorable*, $t(50) = 13.89$, $p < .001$).

Received May 22, 2017
Revision received January 8, 2018
Accepted January 13, 2018 ■