Swayed Away By Numbers: When Consumers Overweight the Review Counts in Their Decisions

Jared Watson, University of Maryland, USA
Anastasiya Pocheptsova, University of Maryland, USA
Michael Trusov, University of Maryland, USA

We investigate how the number of online product reviews affects consumer behavior. We find that most online retailers feature a small review count, and that consumers are more likely to defer purchase or choose the suboptimal alternative when viewing a choice set with a small review count.

[to cite]:

[url]:
http://www.acrwebsite.org/volumes/1019903/volumes/v43/NA-43

[copyright notice]:
This work is copyrighted by The Association for Consumer Research. For permission to copy or use this work in whole or in part, please contact the Copyright Clearance Center at http://www.copyright.com/.
The Effect of Numerical Markers on Consumer Inferences and Decisions
Chair: Jared Watson University of Maryland, USA

Paper #1: The Effect of Reward Quantification on Actual Program Participation
Eleanor Putnam-Farr, Massachusetts Institute of Technology, USA
Jason Riis, University of Pennsylvania, USA

Paper #2: Swayed Away by Numbers: When Consumers Overweight the Review Counts in their Decisions
Jared Watson, University of Maryland, USA
Anastasiya Pocheptsova, University of Maryland, USA
Michael Trusov, University of Maryland, USA

Paper #3: “18% off the Original Price Then Another 12% off” or “12% off Then 18% off”: How Multiple Discounts Influence Consumer Evaluations
Rajesh Bagchi, Virginia Tech, USA
Derick Davis, University of Miami, USA

Paper #4: The Rating Polarity Effect: Overcoming the Surreptitious Influence of Implicit Numerical Associations on Consumer Judgments
Ellie Kyung, Dartmouth College, USA
Manoj Thomas, Cornell University, USA
Aradhna Krishna, University of Michigan, USA

SESSION OVERVIEW
Managers often use numerical markers to motivate and attract consumers. But is this always the best strategy? Imagine a consumer deciding whether to join a rewards program: is she more likely to join one that says “earn up to $125 in rewards” or one that features the message “earn rewards”? How does she use the numerical information in the decision process? Reliance on numerical attributes is a common decision heuristic (Adaval 2013). Consumers use such attributes to answer questions such as “How long?”, “How many?”, or “How much?” to aid in their decision, even when numerical attributes are not the most important information for their decision process.

While recent research on numerosity has advanced our understanding of how various attributes of a number (e.g., magnitude, format, or precision) affect consumer preferences between choice options (Coulter and Coulter 2010, Monga and Bagchi 2012, Zhang and Schwarz 2012), much is yet to be discovered; especially pertaining to how the use of numerical markers in consumer decisions can lead to suboptimal outcomes. This session takes an interdisciplinary approach to the issue, by combining papers with quantitative, behavioral, and managerial perspective to provide an integrative view of the cases where marketers rely on numerical markers to affect behavior in several consumption domains. Across four papers, we find that consumers typically overweight the numerical attributes in their decisions, frequently leading to suboptimal decisions.

In the first paper, Putnam-Farr and Riis investigate how the quantification of rewards motivates program enrollment and participation. In a field experiment with a large wellness provider, they find that a quantified reward (vs. non-quantified) reward messages leads to higher enrollment. Yet, it also leads to an earlier dropout from the program for those who enroll. Next, Watson, Pocheptsova, and Trusov demonstrate a ratings count effect in which consumers’ preference between options systematically changes based on the total number of reviews available on the website. Bagchi and Davis continue to explore the relationship between numerical markers and consumer evaluations by investigating the order of the presentation of multiple discounts. The authors demonstrate that a large-small discount ordering leads to higher product evaluations. Lastly, Kyung, Thomas, and Krishna investigate a rating polarity effect in which consumers encountering an atypical evaluation scale are less sensitive to quality differences between products. The authors find that interference between numerical associations in memory drives this effect.

Taken together, the four papers contribute to the literature by illuminating different consequences of numerical markers on consumers’ inferences and behavior, with an emphasis on how the numerical markers may lead consumers to make suboptimal decisions. With diverse methodology and theoretic lenses applied to the issue, this session should appeal to a broad audience interested in numerosity, inference making, and consumer decision-making, with an emphasis on implications of using numerical markers in marketing actions.

The Effect of Reward Quantification on Actual Program Participation
EXTENDED ABSTRACT
For many programs, ongoing active user participation is an important component. Examples include programs for exercise, diet, frequent buyer rewards, and financial budgeting. For these types of programs, marketers have a dual purpose of both exciting people enough to enroll, but also motivating people to participate. Given that measuring participation outcomes is difficult, but measuring click and enrollment responses is relatively straightforward, most marketers focus on techniques to increase enrollment, with the hopes that participation will follow. This leads to an understandable tendency to focus on attention-getting techniques, such as the maximum potential reward that participants could earn. In this particular set of experiments, we look at a program which offers rewards and consider how the language used to describe these rewards affects both enrollment and ongoing participation.

The tendency to focus attention on the maximum potential reward is a reasonable one. Research suggests that overstatement of benefits can lead to more favorable ratings of product quality than understated reviews (Olshavsky & Miller, 1972). And generally, higher incentives are more attractive to lower incentives for motivating participation (Locke, Latham, & Erez, 1988). However, this emphasis on the maximum potential reward may cause people to focus on that reward, particularly if that potential is expressed in the form of a number. The mere presence of a number, regardless of its validity as a target, can often act as an anchor for expectations from which people insufficiently adjust (Tversky & Kahneman, 1974) which is usually employed when people are asked to judge the probability that an object or event A belongs to class or process B; (ii. Research on lotteries and gambling has shown that demand for lottery tickets is driven by the maximum potential payout, rather than a calculation of expected return (Forrest, Simmons, & Chesters, 2002) and that gamblers often focus on the amount they could win as an anchor (Lichtenstein and Slovic 1971).

If people do use this maximum potential reward as an anchor, it could have significant effects on their eventual satisfaction with the program. On the positive side, the ongoing expectation creates an initial frame of reference that is then updated using information about performance relative to the expectation (Oliver, 1980), so the high target might generate sticky positive expectations. On the other
hand, product performance evaluations can depend more on discrepancies relative to expectations than on actual performance (Weaver and Brickman 1974), so if people fail to perform at the expected level, this could have a negative impact on their satisfaction. Similarly, research on goal performance suggests that people update their opinions about the importance of the goal based on performance towards that goal. In some cases, dissatisfaction with performance on a goal leads people to try to improve their performance, suggesting they will push to achieve the high target (Fishbahar, Dhar, & Zhang, 2006) keeping in shape. However, if the lack of performance is perceived as a personal failure, it can result in decreased commitment to a goal (Soman & Cheema, 2004), particularly if they feel the target is unachievable.

In an initial online field experiment (N=8,918), we tested different email recruitment messages in partnership with a large wellness provider, and measured clicks, enrollment, and ongoing participation. We randomly assigned potential participants to receive either a quantified recruitment email (“earn up to $125 in rewards”) or a non-quantified recruitment email (“earn rewards”) and found that those who received the quantified email were much more likely to click the enrollment link (33% versus 24%, p<.001) and somewhat more likely to enroll in the program (10.02% versus 8.83%, p=.05). This supports the idea that the high target attracts positive attention and motivates people to enroll. However, the average duration of participation for those who enrolled from the quantified message was lower (38 days versus 41 days out of the 60 days measured, p=.05) and more of them dropped out both on the first day and during the first two weeks, despite earning just as much in rewards as those from the non-quantified condition.

In a follow up study on Amazon’s Mechanical Turk, we tested several different potential recruitment messages which varied the language used to describe the potential reward amount (“up to $50”, “up to $20”, “$20 or more”, or no specific target given). We found that people adopt numbers in recruitment messages as personal targets, and are much more likely to be dissatisfied with the program if they do not meet those targets. People who are told they can earn “up to $50” for participation expect to earn substantially more than those who were told they could earn “$20 or more” ($40 vs $30, t=-4.7, p=.001), even though the conditions for earning rewards were the same in both conditions, and up to $50 has a firm upper bound, which should limit the potential reward.

These results help us understand how people shape their expectations about goal formation and progress and thus how marketers can target messages more appropriately to encourage ongoing participation. Goal formation is not necessarily an explicit consideration of what would be a reasonable target, but may be influenced by the language used to describe the program generally. Marketers should be aware of this issue when describing benefits to potential participants, in order to avoid frustrating their customers if these targets are unlikely to be reached.

**Swayed by the Numbers: The Consequences of Displaying Review Counts in Purchase Decisions**

**EXTENDED ABSTRACT**

Imagine a consumer shopping online for a new blender. She narrows her choice set down two options that vary on product rating (e.g., 3.5 vs 3.2). Would her choice be different if she was shopping on a popular website that has a large number of consumer reviews as opposed to a smaller retailer with a small number of reviews?

While previous literature has examined the influence of online reviews on evaluations of a single product (Chevalier and Mayzlin 2006), we investigate how consumer choice between multiple products is affected by the review counts. Consumers have been shown to incorporate others’ purchase decisions into their own (Murray 1991; Bannerjee 1992; Brown and Reingen 1987) which helps reduce uncertainty in their choices (Roselius 1971). Thus, the total review count on the website gives a signal to the consumer of others’ purchase decisions and evaluations, increasing decision confidence as the review count increases. In contrast, when the total review count on the retail website is low consumers’ confidence is decreased.

We propose that ironically consumers’ low confidence when shopping on retail websites with low review counts will lead them to overweight the difference between review counts of the choice options. This happens because consumers are relying on relative differences rather than absolute ones as a simplifying decision heuristic (Fox and Levav 2000). This, we argue, can result in consumers making suboptimal decisions, i.e., choosing lower-quality options when they feature higher review counts, when shopping from retailers with low (vs. high) review counts. Furthermore, we argue that consumers are more likely to defer when the choice set features small review counts (compared to large or absent review counts), yet a majority of online retailers feature small ratings counts, potentially causing adverse effects to their sales.

Study 1 examined the current practices regarding online reviews disclosure. 91 websites from three of the highest grossing online product categories (electronics and appliances, clothing and accessories, and computer hardware; NRF E-Commerce Sales 2010) were analyzed and coded by two independent raters. Most of the websites analyzed disclosed review counts (76%). Of these, 62% featured a small review count, suggesting potential adverse influences to sales based on our prediction of consumer behavior described above.

Study 2 (N=183) tested these predictions. Participants were randomly assigned to one of three review count conditions (large, small, or absent) in a between-subjects design and were asked to imagine they were shopping for a blender. Two blenders were described by a short list of attributes, varying on quality (the number of stars given by other customers) and review counts. While the absolute difference in review counts between the options was kept constant across conditions, the relative difference varied by adding equal numbers of reviews to both options. Next, participants were asked “If you were shopping for a blender in this price range, would you: Purchase Option A [higher product rating, smaller review count], Purchase Option B [lower product rating, larger review count], or Delay purchase to look for other options?”. Finally, participants reported how much different attributes (including the number of reviews) factored into their decision process.

Results confirmed our hypotheses. Participants were significantly more likely to defer to the small review counts condition (Psmall=29%) than the absent (Pabsent=8%; χ²(1,120)=8.514; p<.05) and the large (Plarge=13%; χ²(1,122)=4.856; p<.05) review counts conditions. Of those who chose a product option, participants were significantly more likely to choose the lower-rated product (Option B) in the small ratings count condition (Psmall=64%) than the absent (Pabsent=29%; χ²(1,98)=12.431; p<.001) and the large (Plarge=29%; χ²(1,97)=11.953; p<.001). We further find that the participants were significantly more likely to judge the number of reviews as more important factor in their decision process in the small review counts condition (Msmall=5.59) compared to the absent (Mabsent=5.18; t(180)=1.798; p<.05) or large (Mlarge=5.10; t(180)=2.186; p<.05) conditions. In the next two studies, we demonstrated the generalizability of this finding in other product categories (Study 3a), and via consequential consumer choice (Study 3b).

Studies 4 and 5 examined the underlying process: consumer confidence when shopping on the websites featuring low review
counts. In study 4 (N=364) participants were randomly assigned to one of two review counts conditions (large vs. small), and made a choice between two lounge chairs with a process similar to study 2. After indicating relative preference, participants were asked to “indicate the level of confidence you felt in making the choice” on a 7-point scale. Once again, differences in preference emerged between ratings count conditions, with participants preferring Option B (lower rating, higher review count) more when the ratings count was small (M large-small = 2.49, M small-large = 3.04; F(1,369) = 8.526, p<.01). This effect was mediated by consumers’ confidence (β=−.09; 95% CI: -1.19 to -.03). Study 5 (N=433) was designed to test how the quality of the choice set options moderated this process. We hypothesized that lower product ratings would decrease choice confidence, increasing reliance on review volumes, while higher ratings would increase confidence, mitigating the need to rely on review counts. The study used a 2(volume: high, low) by 3 (product ratings: low, medium, high), and showed that previously reported effect was attenuated in the high product rating condition (index of moderation: β=−0.5; 95% CI: −0.14 to −0.06), with consumers exhibiting high confidence when the choice set featured highly-rated options (e.g., 4.5 vs 4.2) mitigating the effect of review count on preference between choice options.

Taken together, the results of the six studies demonstrate that the review counts disclosed by an online retailer systematically affect consumer preference and purchase intentions. Consumers shopping on the websites with low review counts overweight relative differences in review counts between choice options, leading them into potentially suboptimal decisions, and increasing choice deferral. These effects were attenuated with choice sets featuring highly-rated products, increasing consumers’ confidence and decreasing reliance on the number of reviews in their decision process. These results have significant implications for the online retail environment, as Study 1 showed that majority of online retailers feature a small review count. Our results suggest that these retailers would be better off not revealing review count information, or investing resources in enticing their customers to leave online reviews.

“18% off the Original Price Then Another 12% off” or “12% off Then 18% off”: How Multiple Discounts Influence Consumer Evaluations

EXTENDED ABSTRACT

Imagine a store offers two discounts for a $100 leather jacket: Take 18% off the original price then an additional 12% off at checkout. Here the smaller discount follows the larger one (large-small ordering). What if the discounts were reversed? 12% then 18%; i.e., small-large ordering. How would this affect consumer evaluations (e.g., purchase likelihood, deal evaluations)? Of surprise to many consumers, both orderings yield the same total discount of 27.84%. We show that large-small ordering elicits more favorable evaluations. Here both percentages were presented using a simultaneous mode; but multiple percentages can also be encountered sequentially—at different points in time, one discount at the store entrance, the other at check out. With sequential mode, we predict a reversal—small-large ordering will be evaluated more favorably. Interestingly, if consumers actually calculated the total discount, order or mode should not affect judgments, as the total discount is 27.84% regardless. However, consumers are cognitive misers, and seldom perform the associated computations.

We draw from two research streams—anchoring and order effects—to support our theory. When two percentages are presented simultaneously, we expect consumers to anchor on the first percentage and form a judgment about the total discount—consistent with literature on anchoring (Epley and Gilovich 2010). Consequently, large-small ordering (18%-12%) versus small-large (12%-18%) will lead to the inference that the overall discount is larger. However, with sequential mode (separated by space and/or time), consumers are unaware of the second percentage when encountering the first one. Here too, we expect anchoring to occur, and we expect consumers to make initial judgments based on the first percentage. However, when encountering the second (unexpected) percentage they will use the initial percentage to evaluate the second percentage. If this second percentage is larger (vs. smaller), it makes the deal seem better. This would also lead to an order effect but opposite of that observed for the simultaneous condition; a large-small order (18%-12%) will lead to the inference that the overall discount is smaller than a small-large order (12%-18%). This is because the second discount will seem larger (smaller) when the smaller (larger) discount is provided first, and lead to the perception that the overall deal is better (not as good).

We support our theorizing in three experiments. In experiment 1 we vary order but only use a simultaneous mode. As in the introductory example, participants (N = 61) were considering purchasing a $100 jacket, which was on sale. We manipulated discount order to be large-small (e.g., 18% followed by 12%) or small-large (12%-18%), thus resulting in final price of $72.16 (and a 27.84% discount). An ANOVA with deal perceptions elicited a main effect of order (F(1,59) = 6.29, p < .02); the deal was adjudged better with a large-small (vs. a small-large) discount ordering (M large-small = 5.36 vs. M small-large = 4.67). Thus, with simultaneous mode, individuals anchored on the initial percentage and thought the deal was better when the larger percentage came first.

Experiment 2 was designed to test the effect of order and presentation mode (simultaneous vs. sequential) together. Participants learned they were planning a 4-day all-inclusive vacation (including hotel, food, and activities), and find a package for $590 (retail). The package was discounted. There were two discounts (11% and 4%), which were either presented on the same page as the scenario (simultaneous mode) or on different screens (one with the scenario, and the other one on the subsequent page; sequential mode). The order of the discounts was also varied, 11% or 4% presented first (large-small or small-large, respectively). Thus, a 2 Presentation (Simultaneous vs. Sequential) x 2 Order (Large-small vs. Small-large) between-subjects factorial design was used.

Participants (N= 217) indicated purchase likelihood and deal perceptions. Independent ANOVAs with both dependent measures elicited only the predicted two-way interactions (F(1, 213) = 13.95, p < .001, and F(1, 213) = 7.96, p < .01, respectively). Participants in simultaneous conditions wanted to purchase the package more with large-small ordering (M large-small = 5.41 vs. M small-large = 4.66; p < .01), but in sequential conditions, the small-large ordering discount was favored (M large-small = 4.75 vs. M small-large = 5.42; p < .05). A similar pattern of means emerged for deal perceptions (Simultaneous: M large-small = 5.62 vs. M small-large = 5.21; p < .06; Sequential: M large-small = 5.38 vs. M small-large = 5.82; p < .05). Furthermore, deal perceptions mediated the effect of our independent variables on purchase likelihood.

Experiment 3 replicates our findings in a laptop purchase context with percentage increases (as opposed to decreases in the earlier studies). Participants (N = 189) read about battery life improvements to the model they were interested in. The presentation of the improvements in percentage increase in battery life constituted our manipulations in a 2 (Presentation: Sequential vs. Simultaneous) x 2 (Order: Larger increase first vs. Smaller increase first) between-subjects design. In the simultaneous (sequential) mode, both percentage increases were on the same (different) screens. One improvement
was a software update (improved battery life by 13%), and the additional improvement was an upgraded hard drive (improved battery life by an additional 8%). The order of improvements was also varied (13%-8% vs. 8%-13%). An ANOVA with upgrade perceptions yielded a significant two-way interaction of mode and order (F(1, 185) = 12.29, p < .001). In simultaneous mode, large-small ordering was evaluated more positively ($M_{\text{large-first}} = 5.04$ vs. $M_{\text{small-first}} = 4.57$; contrast $p < .04$), whereas the opposite was true in sequential mode ($M_{\text{large-first}} = 4.41$ vs. $M_{\text{small-first}} = 5.01$; contrast $p < .01$).

Thus, we show how consumers evaluate multiple percentage changes, and contribute to literatures on numeracy, marketing, finance, and economics. In all of these areas computations play an important role (e.g., evaluating loans with compound interests or assessing stock price changes across quarters). In economics too, normative and prescriptive theories are developed based on preference for multiple lotteries using percentage-based probabilities.

**The Rating Polarity Effect: Overcoming the Surrupitious Influence of Implicit Numerical Associations on Consumer Judgments**

EXTENDED ABSTRACT

Numeric ratings are frequently used to describe products and are an important input for consumer judgments. However, previous research has not examined how culturally determined implicit numerical associations might influence such judgments depending on the rating format employed. Numeric ratings can be presented using a bigger-is-better ($1=\text{bad}, 5=\text{good}$) or smaller-is-better ($1=\text{good}, 5=\text{bad}$) format with reversed rating poles. The format that consumers find typical will depend on their cultural context. Consumers might encounter numeric rating systems with atypical formats when traveling abroad, or even in their own country (e.g. NIH grant proposals in the U.S.). Our research is the first to examine how rating format influences judgments, demonstrating a rating polarity effect: Consumer evaluations are less sensitive to differences in product quality when using rating formats with a numerical association atypical for their culture. Decades of research on memory and judgments suggest that two different types of memory processes influence our everyday judgments: rules stored in explicit memory and associations stored in implicit memory that spontaneously, and sometimes surreptitiously, influences judgments (Graf and Schacter 1987; Schacter 1987). Previous research in numerical cognition suggests that people can form implicit associations with numbers based on their cultural context (Dehaene, Bossini, and Giraux 1993). We posit that the rating polarity effect is caused by interference from typical, implicit numerical associations when applying atypical, explicit rules and that mindsets that increase reliance on implicit associations versus explicit rules can mitigate the effect of interference on consumer product evaluations.

Our experiments use an interference paradigm (Jacoby 1991) where measuring the extent of interference that comes from automatic versus intentional uses of memory requires an experimental paradigm that compares outcomes where the implicit association and explicit rule coincide versus where they collide. Thus we compare participant product evaluations when using typical rating polarity, where the implicit numerical association (bigger-is-better) and explicit rule (bigger-is-better) coincide, to product evaluations when using atypical rating polarity, where the implicit numerical association (bigger-is-better) and explicit rule (smaller-is-better) collide. The difference in product evaluations between participants using the bigger-is-better versus smaller-is-better rating polarity reflects the extent of interference from the implicit numerical association when attempting to apply an explicit rule. Participants are presented with five brands in three product categories (water, margarine, toothpaste). For each brand, they see the brand name, photograph, tagline, and quality rating (1, 2, 3, 4, or 5). In the bigger-is-better condition, 1 = inadequate and 5 = very good. In the smaller-is-better condition, 1 = very good and 5 = inadequate. Participants then indicated their purchase intentions for each product. The presence of interference is indicated by a significant rating polarity x product quality interaction.

Experiment 1 demonstrates how interference between numerical associations can affect consumer product evaluations. American participants are presented products with quality ratings using rating polarity with either a bigger-is-better (typical) or a smaller-is-better (atypical) numerical association. Repeated measures logistic regression revealed a significant rating polarity x product quality interaction for both purchase intent ($\beta = -.36, p < .01$) and willingness-to-pay measures ($\beta = -.25, p < .01$), indicating the presence of interference. American participants were less willing to purchase high quality product and more willing to purchase low quality products when using a smaller-is-better numerical association.

Experiment 2 demonstrates the cultural variability of the effect with Germany participants, where a smaller-is-better numerical association is more typical. Repeated measures logistic regression revealed a significant rating polarity x product quality interaction for both purchase intent ($\beta = -.39, p < .01$) and willingness-to-pay measures ($\beta = -.04, p = .06$), again indicating the presence of interference. German participants’ product evaluations were less sensitive to difference in product quality when using a bigger-is-better numerical association.

If the effect of interference on consumers’ product evaluations is due to interference from the typical numerical association in implicit memory with the atypical numerical association in explicit memory, then a mindset that encourages reliance on information in explicit memory should attenuate memory interference. Previous research has demonstrated that judgments and decisions can be characterized along a continuum of ‘why’ and ‘how’ mindsets (Freitas, Gollwitzer, and Trope 2004; Trope and Liberman 2003; Vallacher and Wegner 1987, 1989). When in a ‘why’ mindset, people focus on higher-level thinking—achieving an outcome and why a task is done. This is in contrast to a ‘how’ mindset, which increases lower-level thinking, focused on process and how a task is done. We propose that a ‘how’ mindset can reduce interference from implicit associations when applying explicit rules.

In experiment 3, we measured participants’ propensity to adopt a more ‘how’ versus ‘why’ mindset using the Behavioral Identification Form (Vallacher and Wegner 1989; scores range from 0-24) after they completed the same product evaluation task. The three-way interaction between rating polarity, quality level, and mindset was significant ($\beta = -.024, p = .01$). Using a series of 12 spotlight regressions (Spiller et al. 2013), the coefficient for the two way interaction between rating polarity and quality level (representing interference) was significant for all participants except for those with the most ‘how’ oriented mindsets (Johnson-Neyman value between BIF values of 6 and 8).

In Experiment 4, we manipulated mindset. Participants completed a task to prime a ‘how’ versus ‘why’ mindset (Freitas, Gollwitzer, and Trope 2004) before making product evaluations. The three-way rating polarity, quality level, and mindset interaction was significant ($\beta = .20, p = .02$). The coefficient for the quality level x rating polarity interaction was statistically significant for participants in a ‘why’ mindset ($\beta = -.31, p < .01$), but the magnitude of the coefficient was smaller and marginally significant for participants in a
‘how’ mindset (β = -.10, p = .09). Thus a ‘how’ mindset attenuated the effect of memory interference on product evaluations.

Our research demonstrates that interference between implicit and explicit numerical associations in memory can influence consumer product evaluations and that this effect depends on cultural context. Most importantly, our research shows that mindsets play an important role in reducing interference between implicit associations and explicit rules.

REFERENCES