Exploratory Study on Anchoring: Fake Vote Counts in Consumer Reviews Affect Judgments of Information Quality

Makoto Nakayama[1] and Yun Wan[2]

[1] DePaul University, College of Computing and Digital Media, Chicago, IL, USA, mnakayama@cdm.depaul.edu
[2] University of Houston-Victoria, School of Arts and Sciences, Victoria, TX, USA, wany@uhv.edu

Abstract

Popular products in online stores often have overwhelming numbers of reviews. To help consumers identity quality reviews, (Site 1) crowdsources this decision by asking consumers to vote on the helpfulness of reviews. Many studies assume these votes reflect the information quality of a review, but this does not account for the influence of fake votes. This study investigates whether fake votes influence judgments of review information quality. From controlled questionnaires given to 294 consumers, we found that fake vote counts could completely reverse judgments of information quality in both directions. Specifically, fake votes affected consumers' processes of purchase decision-making and product research. Overall, fake votes changed both review judgments and purchase behaviors.

Keywords: Consumer reviews, Fake votes, Review information quality (IQ), Anchoring, Product learning, Purchase decision-making

1   Introduction

Online consumer reviews (*online reviews* or *reviews* hereafter) offer great value to customers [52]. About 70% of Americans consult reviews before making a purchase decision [4] and 95% of U.K. consumers *always* or *sometimes* consult reviews before purchasing online or in stores [66]. These statistics show that online reviews are critical in consumers' purchase process.

Because reviews are important, review manipulation is becoming popular. A recent Gartner report [77] noted that 2% to 6% of ratings and reviews are deceptive and that an industry of fake reviewers is emerging with a price range of $1 to $200 per review depending on the products. Other algorithms and techniques have estimated that about 2% to 10.3% online reviews are fake [33], [45].

While assessing the content of a review may tell us if it is fake, detecting fake review ratings is much more challenging because a rating does not leave any traces of author authenticity. Because a manipulated rating gives a false signal to the information quality (IQ) of a review, it is important to understand the extent of influences by rating manipulations.

Some websites give users the ability to rate the quality of reviews. Amazon.com (Site 1), for instance, allows consumers to vote on whether a review is helpful or not (e.g., *12 of 20 people found the following review helpful*) as shown in Figure 1. These helpfulness votes (H-Votes hereafter) and their ratios (the total YES votes divided by the total votes) are used to sort reviews in order of review helpfulness. In addition, Amazon.com (Site 1) prominently displays the two most helpful positive and negative reviews for each product.

Figure 1: Prominently displayed most helpful favorable (positive) and critical (negative) reviews (Site 1)

Previous studies on consumer reviews focused on the relationship between content characteristics and content authenticity (e.g., [32], [53]) or between content characteristics and H-Vote ratios (e.g., [41], [52]). Yet, these studies regard H-Votes as authentic. Some studies assumed consumers used their own judgment to assess the review information quality (IQ) regardless of the number of H-Votes. However, the *anchoring effect* [80] can affect how consumers interpret existing H-Votes. That is, consumers may be more likely to rate a review as *helpful* when they see a higher H-Vote ratio than a lower one. If so, the existing H-Vote ratio would influence subsequent votes in a cascading manner. Thus, our first research question is *to what extent does the anchoring effect exists in H-Vote ratios*?

Anchoring may not only influence the moment of judging review IQ. It may also affect the product learning process that consumers typically go through before making a purchase decision. In this paper, we refer to *product learning* as the exploration of product functionality and other aspects, equivalent to the intelligence and design phases in Simon's decision-making process [70]. In contrast, *purchase decision-making* refers to the choice selection and purchase-making phases of his model. Thus, reviews may provide value in long-term learning and in short-term decision-making. For instance, one study found that the average consumer spent 6 hours and 21 minutes learning about a product online, even though they need just 6 minutes to shop at retailer websites [66]. A recent survey notes, "one in two respondents spent 75% of their overall shopping time researching product (sic) as compared to just 21% in 2010" [64]. Thus, our second research question is *to what extent does anchoring affect learning and purchase decision-making*?

Assessing how anchoring affects H-Vote ratios is important for four reasons. First, previous studies rarely examined the anchoring effects on H-Vote ratios. Second, if anchoring is confirmed on H-Vote ratios, how does it influence consumers helpfulness voting behavior? Third, such anchoring may depend on the type of good and review. Fourth, anchoring should be examined for both short- and long-term effects because consumers use reviews to purchase at some later time.

The next section surveys the theoretical background pertinent to our research questions. We then set forth hypotheses, followed by methodology and results. Finally, we discuss the implications of our study, the future research agenda, and the consequences of our conclusions.

2   Theoretical Background

IQ indicators of consumer reviews.   Existing studies examine the quality of reviews by using two approaches. In the first approach, researchers ask which content characteristics indicate the IQ of helpful reviews. The second approach assesses the authenticity of reviews.

To consumers, one of the most visible IQ indicators is the count of YES/NO votes of how many other consumers find a review *helpful*. This count gives an H-Vote ratio, or *summary* IQ indicator. Existing studies use these votes as the most critical IQ judgments from consumers. Assuming the votes are authentic, these studies examine what characteristics highly *helpful* reviews have in common. They have identified three basic IQ indicators along with three circumstantial factors (Table 1).

Table 1: IQ indicators of reviews

Study	Summary IQ Indicator	Basic IQ Indicators			Circumstantial IQ Factors		
		Length	Readability	Substance	Good Type	Valence	Total Votes
Mudambi & Schuff [52]	H-V ratio	X			X	X	X
Korfiatis et al. [41]	H-V ratio	X	X		X	X	X
Cao et al. [14]	H-Votes	X		X		X	X
Pan & Zhang [60]	H-V ratio	X		X	X	X	X
Baek et al. [5]	H-V ratio	X	X		X	X	
Kuan et al. [44]	H-Votes & ratio	X	X		X	X	

The three basic IQ indicators of a review are its length, readability, and meaning or substance. Review length is significantly related to IQ because helpful reviews usually use more words to describe the product. Quality reviews are also easy to read: by combining four readability indices to assess the complexity of texts and ease of reading, Korfiatis et al. [41] found that helpful reviews were more readable than other reviews. A review's substance or meaning also influences the review IQ, though determining "the exact meaning of text is extremely difficult and often subjective" [14] p. 515. Using latent semantic analysis (LSA), Cao et al. [13] found that the variety, frequency, and pattern of words (i.e., concepts) used in a review suggest its substance. In addition, Pan and Zhang [60] reported that the innovativeness of a reviewer affected the review IQ, with moderate innovativeness being better than extreme. Reviewer credibility and reputation have also been reported to be factors in review IQ [5], [44].

In addition to the three basic IQ indicators, circumstantial IQ factors such as goods type, valence, and total votes also influence the perception of review IQ. The goods type classifies a product based on whether it is feasible to evaluate its quality before or after using it. Some products have tangible characteristics that can be easily evaluated before purchase (e.g., electronics), while others have subjective characteristics among users (e.g., music, wine). Some products claim to have characteristics that can only be evaluated after long-term usage (e.g., vitamin supplements) or in special circumstances (insurance). Valence can also influence review IQ. A product review with a real name and/or verified purchase is usually considered to have higher IQ than similar reviews without those attributes. Finally, the total vote count affects IQ because it is related to the number of shoppers who have read a review, making it a circumstantial indicator of the review's popularity, which directly relates to its quality.

The second approach to review IQ is to question if a review is fake or not. Much research has been done to optimize spam-detection algorithms and models (e.g., [38], [46], [48]). However, detecting fake reviews requires detailed data on the review's posting date, the reviewer's ID, and the textual characteristics of reviews. Average consumers have no easy way to tell which reviews are fake, even when reviews have some suspicious characteristics (e.g., multiple reviews on a product from the same reviewer, use of extreme language, unfair views on product features).

In theory, consumers have they need to judge the IQ of a review. They can consider its length, readability, and substance: the three basic IQ indicators. However, this judgment does not seem simple because the basic IQ indicators that make a review *helpful* vary by the good type and review valence. Also, consumers can see how others judged the same review, which is the summary IQ indicator.

Consumer decision-making. During the last few decades, consumer decision-making research has grown significantly, producing a number of new decision theories and models [9], [73] that depend on consumer profiles (e.g., demographics), product categories (e.g., utilitarian vs. non-utilitarian goods), and purchase types (e.g., in-store vs. online purchase). Smallman et al. [74] summarizes the consumer decision-making paradigms: from prescriptive, analytical decision-making to bounded rationality, adaptive decision-making, and more recent pragmatic and naturalistic decision-making. We limit the scope of the present literature review to the issues associated with online consumers, primarily regarding online purchase decision-making.

Several studies have proposed models of online consumer decision-making. Two studies [39], [88] offer consumer motivation models that incorporate the technology acceptance model [18], the theory of reasoned action [26], and/or

Makoto Nakayama
Yun Wan

the theory of planned behavior [2]. These models do not include variables on information-seeking behaviors, but two others do: In the first, Kohli et al. [40] apply Simon's [70], [72] decision-making model. In the second, Darley et al. [16] present an extended model based on Engel et al. [22], [23]. These two models remind us that consumer decision-making is a *process* in which information searching and learning about products play indispensable roles. However, these four studies do not assess how anchoring might affect reviews.

With the emergence of online searching and purchasing, the cognitive efforts of consumers are now supplemented by "electronic decision aids" [7] such as H-Votes, star-rating summaries, and prominent displays of the most helpful reviews. However, the effectiveness of these electronic decision aids is somewhat inconclusive, according to empirical studies [29]. A common decision aid is the product recommendation system, now used widely at retailer websites. For example, (Site 1) displays recommended products-based on a customer's profile, purchase history, and other factors-using sophisticated algorithms such as collaborative filtering and cluster modeling [49]. Such tools go hand in hand with the customers' process of learning the true quality of a product. This leads some research to focus on how customers learn (e.g., [69], [90]). Many electronic decision aids include review quality or helpfulness in their recommendation factors. However, helpfulness votes can be manipulated, and these fake votes can add up to change the overall opinion [65], [89]. The influence of fake IQ indicators has been understudied, but they are becoming increasingly importance in consumer learning, so they should be examined in the context of reviews.

Anchoring-and-adjustment heuristic. Judging the IQ of a review is not a simple, objective task because of the varying-and often frequently complex and subjective-product characteristics and uncertainty in review authenticity, as noted earlier. Under these circumstances, consumers often rely on other consumers' judgments as anchors to start their evaluation and adjust their own judgments as needed. This phenomenon is known as the anchoring-and-adjustment heuristic [24], [82]. According to Tversky and Kahneman [80], individuals make estimates by starting from an initial value and adjusting it to reach their final answer. Offered different starting points, individuals anchor their final estimations toward those values, leading to biased outcomes.

Anchoring is well studied and often used in marketing, especially in pricing tactics [3], [30], [68], [81] such as price comparison, everyday low pricing, and reference price anchoring. This type of anchoring is commonly known as numerical anchoring, which is defined as "the assimilation of a quantitative estimate towards a previously presented number" [67]. Numerical anchoring has been examined, producing various findings: numerical anchoring is stronger under time pressure [43], subliminal anchoring exists [67], and anchoring depends on access to other considerations [27]. Also, when anchors are placed outside the plausible value range, their effects diminish [83], [84]. Recent studies [10], [27], [84] also consider the elaboration likelihood model (ELM) [62], [63] and how anchoring affects decision-making in situations that require higher or lower cognitive loads. Numerical anchoring affects decision-making both in relatively thoughtful, high-elaboration processes and in relatively non-thoughtful, low-elaboration processes. Interestingly, when asked to assess the price a second time, individuals in high-elaboration processes adjusted less than those in low-elaboration processes [83].

Most research has focused on numerical anchoring in pricing policies, and few studies have explored how anchoring can influence judgment of review IQ. In consumer reviews, some IQ indicators are numbers, while the others are not numbers. The summary IQ indicator, in particular, is a ratio of two numbers: YES votes and total votes. Thus, from an anchoring perspective, consumer reviews pose a unique challenge to investigate the influence of anchoring.

Anchors, the data giving an initial reference point for a judgment, are likely to influence consumers. Similarly, summary IQ indicators (H-Vote ratios) may act as anchors, not only in individuals' judgments of review helpfulness, but also in their learning process of the product (as delayed anchoring effect). This leads us to our research objective of studying how anchoring from fake summary IQ indicators influence consumer reviews at two levels (Figure 2): First, fake summary IQ indicators may change how consumers vote on review helpfulness at the YES/NO level. Second, they may influence consumers' short-term purchase decisions and ongoing product learning, ultimately changing purchase behaviors.
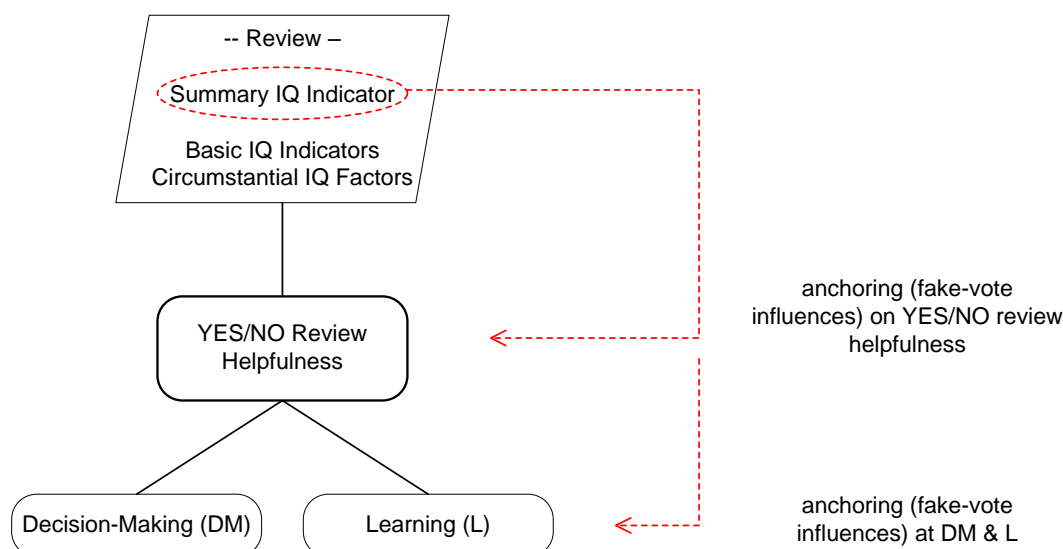
Makoto Nakayama
Yun Wan

Figure 2: Basic research questions

3   Hypotheses

These study addresses two basic questions.   First, does anchoring change consumer IQ judgments when consumers vote on the helpfulness of a review? Second, how does anchoring influence the two latent foci (learning and decision-making) differently as consumers rate review IQ? While more than one indicator affects review IQ, (Site 1) *most helpful* reviews (Figure 1) enables us to control the three basic IQ indicators and total votes. Because these *most helpful* reviews are given for a specific product and review valence, we can examine our research questions for specific review valences and product types (Figure 3). In the following, we first define product types and then elaborate our hypotheses that address these two basic questions.
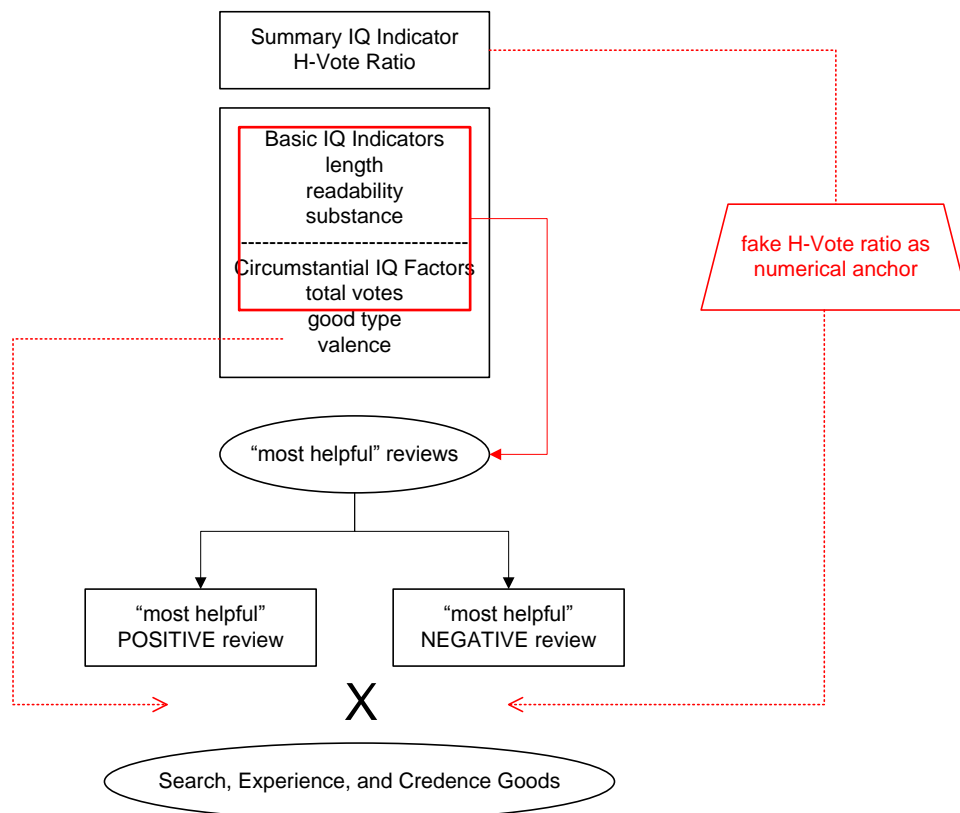

Figure 3: Examining anchoring effects with different review IQ indicators

3.1 Search Experience and Credence (SEC) Goods and Decision Anchoring

Consumers evaluate review IQ differently based on goods type. While there are several ways to categorize goods, our research categorizes them into SEC goods. Search goods are those whose quality can be evaluated prior to purchase or use. Experience goods are those whose quality can be examined after use [55]. Credence goods are those whose qualities *cannot be evaluated in normal use* even long after their purchase [15]. We use these three types for the following reasons.

Extant studies [31], [35], [52], [54] on online consumer behaviors and reviews frequently select products based on SEC categories. The searchable characteristics of search goods are particularly relevant in our study because reviews of search goods should be based on objective claims about their tangible attributes [52]. If the claims made by a review are objective and conceivably credible, there may be less room for manipulation. In contrast, the quality of experience goods depends on both objective and subjective experiences [86]. Thus, advertisement more strongly affects the perceived quality of experience goods than search goods [56]. The quality of credence goods is even more difficult to evaluate and thus may be even more easily influenced by advertisement. Because helpfulness voting is similar to advertisement in how it influences consumers' evaluation of review quality, we use this classification method to examine how anchoring affects the three goods types.

3.2 Does Anchoring Exist in YES/NO Votes?

For a search good, consumers appreciate a review more when it specifically addresses the tangible product characteristics: for example, *This printer took just 1 minute to print 80 pages, and the prices of similar printers are at least $50 higher than this printer*. In contrast, credence goods have few tangible benefits that can be easily evaluated, so consumers might appreciate reviews that addresses reasons why they should avoid buying the product: for example, *I had an allergy to this vitamin*. Consumers also adopt a decision strategy that varies depending on the goods type. For example, when a consumer evaluates search products, it is easier to use a compensatory strategy; when evaluating a credence good, a non-compensatory strategy might be more effective [20], [71], [79]. Table 2 shows the preferred content and decision strategy for each goods type.

Table 2: Review content orientation by SEC category

SEC Category	Attribute Characteristics	Positive Reviews	Negative Reviews
Search Goods	searchable, measurable, tangible	*compensatory* assessment with searchable attributes	value judgment on functions, price comparisons to alternatives
Experience Goods	non-searchable, subjective, intangible	*Compensatory* assessment with subjective attributes	*non-compensatory* assessment using elimination-by-aspects (EBA)
Credence Goods	non-searchable, hard to define benefits	rational assessment is difficult with indefinable benefits	*non-compensatory* assessment using elimination-by-aspects (EBA)

A *compensatory strategy* uses a weighted-additive rule for the benefits of search goods. In this approach, a consumer first assigns a weight to each attribute, finds a total score for each product based on the sum of the weighted attribute scores, and then selects the product that has the highest total score [10]. This strategy is easy to use for search goods whose attributes are tangible. It can also be applied to experience goods because consumers can add up the qualitative positive experiences of other consumers. Note that most online reviews are not well structured: these reviews are considered an informal form of knowledge sharing [61]. Thus, consumers must mentally construct a *compensatory* assessment of attributes from the review. The content of such a review is then subject to anchoring from the summary IQ indicator.

A *non-compensatory strategy* utilizes elimination-by-aspects (EBA): If a product fails to satisfy particular requirements, it is dropped from consideration. This process is repeated until one product remains [79]. Consumers can use EBA even without complete information about products' features and functions. EBA is especially useful in evaluating the quality of negative reviews. When it comes to negative reviews, anchoring may sway review judgments more for experience and credence goods than for search goods. Compensatory and non-compensatory strategies are applicable and valid for consumer decision making, remaining robust even in recent empirical studies [37], [58], [59], [78].

Overall, when consumers read positive reviews of products whose attributes and benefits are tangible, these review can give them a list of clear reasons to buy the product or not. Similarly, when consumers read a negative review for products whose attributes or benefits are unclear but whose descriptions in the review are *damaging* can give them a clear signal to stay away from the product. Thus, we hypothesize:

H1a: For search goods with easily verifiable tangible benefits, consumers use a compensatory strategy, and manipulation of summary IQ indicators influences positive reviews.

H1b: For experience and credence goods with non-verifiable tangible benefits, consumers use a non-compensatory strategy, and manipulation of summary IQ indicators influences negative reviews.

3.3   Do the Two Hidden Foci of Review Helpfulness Exist?

Consumers buy a product based on two sub-processes: learning and purchase decision-making [76].

Learning is defined as "the vicarious acquisition of cognitive competencies that consumers apply to purchase and consumption decision-making" [13]. Decision-making is the next logical step of product selection, in which consumers make a purchase decision. Consumer learning is a central construct used in consumer behavior models [36], but psychological decision research tends to neglect the influence of learning [8], [9]. For example, it was not until recently that some studies examined the influence of learning in online environments (e.g., [13], [89], [90]).

Reviews could help consumers in both learning and decision-making. These two foci are intimately related-the same review may help in completely different ways regarding these two foci, according to different consumers or even the same consumer in different sub-processes depending on context-so the consumer must decide on a review's helpfulness in both of these aspects [57].

Additionally, the characteristics of SEC goods may affect consumers' perceptions of helpfulness differently in learning and decision-making [17]. Search goods have attributes and functions that are easier to verify, so consumers of these goods-whether they are in the learning or decision-making sub-process- could easily grasp the review content according to their needs. In contrast, credence goods have attributes that are very difficult to verify, so consumers engaging in either learning or decision-making would be challenged in clearly differentiating the helpfulness of either aspect. Thus, we hypothesize the following:

H2a: Consumers will perceive the helpfulness of a given review differently based on differences in their needs related to their learning and decision-making sub-processes.

H2b: Depending on the goods type, consumers experience different levels of difficulty in determining the helpfulness of a review's content in learning and decision-making because of differences in how difficult it is to verify that content. The order of difficulty from easiest to hardest is search goods, experience goods, and credence goods.

3.4   How does Anchoring Affect the Two Hidden Foci Differently?

The next logical question is this: how does manipulation of review helpfulness voting affect the perception of learning and decision-making based on review content? In other words, how does the anchoring effect in helpfulness voting affect these two foci?

The IQ of a review depends on (a) learning values, (b) decision-making values, or (c) both learning and decision-making values. When consumers vote YES or NO on whether a review is helpful, their decisions depend on the weights of the two foci. Some consumers may vote YES based mainly on whether the review helps them decide whether they should buy the product (decision-making). Some vote depending on how much they learn from a review (learning). Other consumers use weighted averages of these criteria for their judgment of review IQ.

The purchase decision-making sub-process is dichotomous (to buy or not to buy), while the learning sub-process is multidimensional and based on key product attributes. Learning could be evaluated with a graded scale such as the Likert scale, similar to valuation using choices with multiple-bound questions [6]. Previous studies [19], [85] indicated that preference distributions are usually different between the dichotomous mode and the multiple-bounded, discrete-choice (MBDC) mode. Thus, we expect that anchoring will have different effects in learning and in decision-making.

The normative decision-making model says that anchoring is used when consumers must engage in conscious value-judgment [11]. Such behavior is likely to occur when consumers make a purchase by adding up the tangible benefits noted in positive reviews of search goods. While reviews for experience and credence goods have more subjective content, their positive reviews can also provide pleasant experiences as scalable units of qualitative benefit, which consumers would add up in their minds.

The learning aspect of positive reviews for search and experience goods affect potentially many cues with respect to anchoring. However, negative reviews of experience and credence goods involve the limited number of cues for decision-making using EBA. Because these goods usually have attributes that are intangible or difficult to verify, the anchoring effect between learning and decision-making foci is likely obfuscating to consumers. Thus, we propose these hypotheses:

H3a: In search and experience goods, manipulations of the IQ indicator, such as helpfulness voting for the most helpful positive review, affects decision-making more than learning.

H3b: In experience and credence goods, manipulations of the IQ indicator, such as helpfulness voting for the most helpful negative review, affects learning and decision-making similarly.

Our theoretical model and hypotheses is summarized in Figure 4 below.
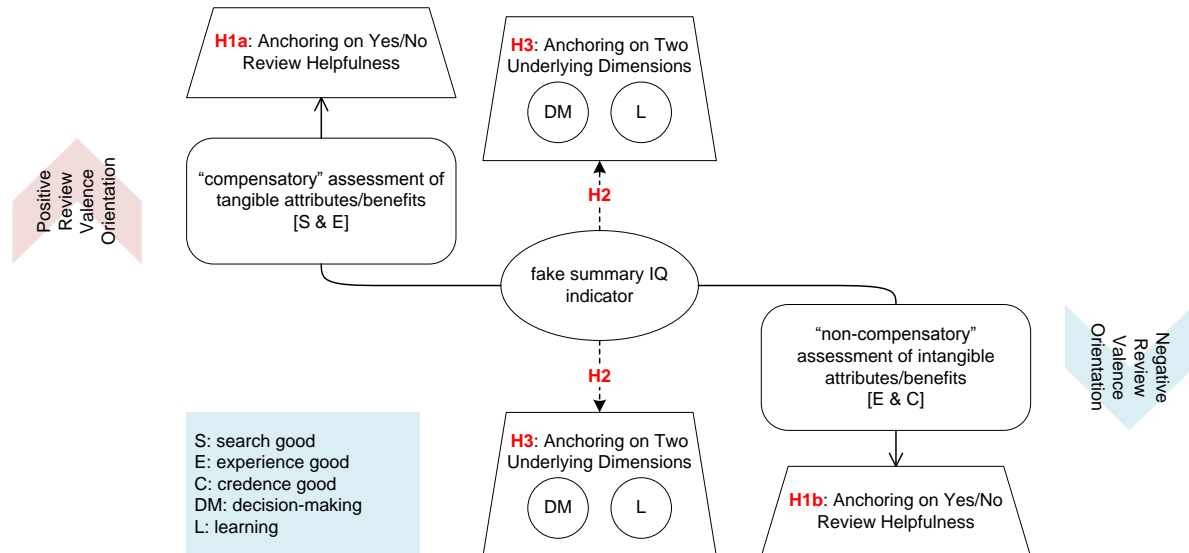

Figure 4: Research model and hypotheses

4 Method

To test the hypotheses, we set up an online survey questionnaire that gives reviews close in appearance to what consumers see on Amazon.com (Site 1). Using the most helpful reviews (as seen in Figure 1), we manipulated the vote ratio of each review in three different ways (as three treatment conditions): a perfect ratio (100% H-Vote ratio), half that ratio (50%), and visible no ratio/votes. Our survey presented each participant with the same three reviews, but with a treatment condition randomly selected from the three (Figure 5). (Site 1), the most helpful positive reviews often have well above a 90% H-Vote ratio, while the most-helpful negative reviews commonly have around 70%. We chose a 50% ratio for the middle condition because it is lower than these typical H-Vote ratios but not as extreme as a 0% ratio.
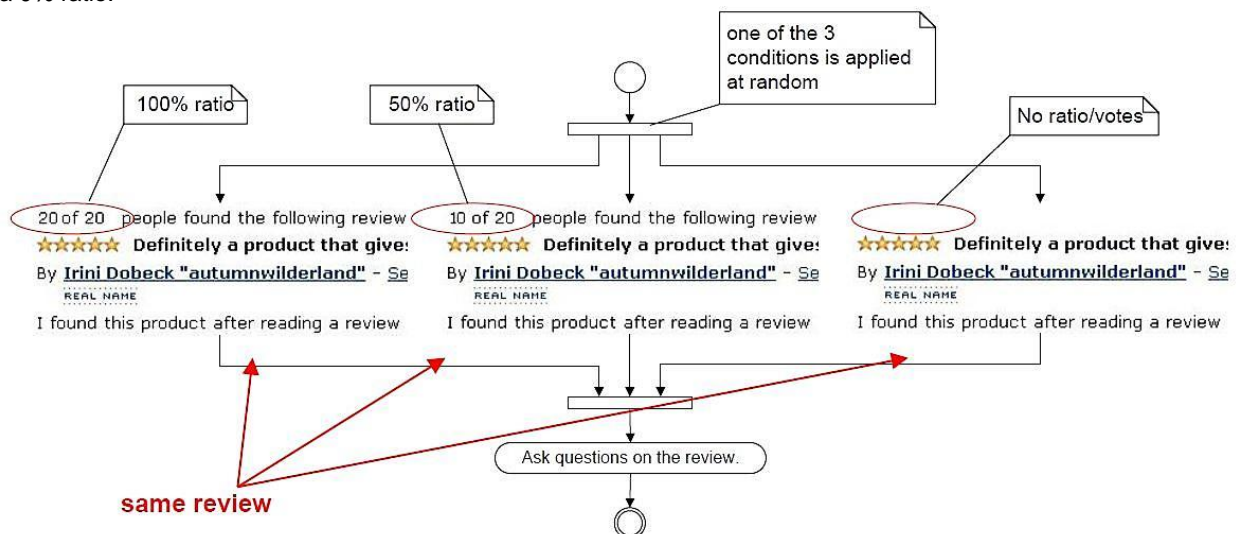
Figure 5: Flow of review IQ evaluation with fake summary IQ indicators (manipulated H-Votes)

After showing a review with a manipulated anchor, we had each participant evaluate the review for (a) its helpfulness with a YES/NO choice, (b) its helpfulness in learning on a 5-point Likert scale, and (c) its helpfulness in decision-making on a 5-point Likert scale (Appendix A). The reviews were manipulated without affecting the font size or

8

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762017000100002

display position of the H-Votes. Prior to administering the questionnaire, we conducted a pilot study to test the validity of the instruments with 108 students from two Midwestern and Southwestern universities in the United States. We used the same survey platform and variable constructs for the current study. In particular, we tested a review on a vitamin supplement to assess whether most participants have adequate knowledge and experience on such a good. The remarks from the pilot study participants confirmed that they were able to follow the survey questionnaire without confusion. The spread of their answers agreed with our expectations as to which levels of numerical anchors should be used.

Selection of products and their reviews. Two products were selected from each category of search goods, experience goods, and credence goods. We randomly picked them from the first page of (Site 1) Best Sellers in the *Printers* (search goods), *Music CDs* (experience goods), and *Vitamins and Supplements* (credence goods) sections in January 2012. The intent of using only two goods is not to investigate exhaustively all kinds of goods, but to dispute that no anchoring effect is seen across all goods. Printers have the general characteristics of search goods, such as hardware specifications and measurable performance characteristics.   Music CDs are considered experience goods because their value depends on personal subjective tastes and is hard to estimate prior to purchase. These two goods are used as search and experience goods in a recent study by Mudambi et al. [52]. Vitamins and supplements are considered classic credence goods [15]: their effect on heath is not apparent even after extended use. For those six goods, we used a pair of the most helpful positive and negative reviews displayed by (Site 1) at the time. Table 3 summarizes the key characteristics of the selected reviews.

Table 3: Characteristics of most helpful consumer reviews used for the study

SEC Category	Product	Review Valence	Star	Words
Search	Canon MP280 Printer	Positive	5	51
		Negative	3	29
	HP LaserJet Pro 1102w	Positive	5	244
		Negative	3	301
Experience	Adele 21 (Music CD)	Positive	5	224
		Negative	3	68
	El Camino (Music CD)	Positive	5	620
		Negative	3	224
Credence	Viviscal Supplement	Positive	5	512
		Negative	1	188
	Nature Way Coconut Oil	Positive	5	204
		Negative	3	216

Data collection. The data were collected from a survey questionnaire in 2012. We solicited volunteers via (Site 2), offering a $5 (Site 1) gift card for completing the questionnaire. We also solicited volunteers at the two universities where we conducted the pilot study, offering the same incentive. Because online questionnaires may collect multiple or automated entries, we screened invalid survey entries with the following criteria: (1) The participants must be 18 years or older and U.S. residents; (2) The IP addresses of participants must be verifiable U.S. IP addresses (no proxy servers); (3) The participant's name, email address, and mailing address must have some indication of validity by matching name and email address as well as name and address via (Site 3); (4) The text comments must be in English; and (5) All the questions are answered fully. In the end, we collected 294 valid entries.

Sample. We drew survey participants from two universities (n = 125) and the general public (n = 169). We used ANOVA (Analysis of Variance) to see if there were any statistical differences in how the two sample groups cast H-Votes. The results of ANOVA (dependent variables: H-Votes for the 6 goods, factor: the two groups) were not significant. The F values and their p values are as follows: $F_1(1, 292) = 0.013$, $p = 0.908$; $F_2(1, 292) = 0.025$, $p = 0.875$; $F_3(1, 292) = 2.355$, $p = 0.126$; $F_4(1, 292) = 0.376$, $p = 0.540$; $F_5(1, 292) = 2.328$, $p = 0.128$; $F_6(1, 292) = 0.045$, $p = 0.832$. Therefore, we regard the two groups as equal for the purposes of this study.

Sample characteristics. Table 4 shows the sample's characteristics. Traditional undergraduate students (age 18-22) made up only 15% of the sample.

Makoto Nakayama
Yun Wan

Table 4: Profile of respondents

Category	Percent of respondents
Age	
18–22	15.0%
23–35	50.0%
36–45	19.0%
46–55	10.9%
56–65	4.4%
66 or above	0.7%
Gender	
Male	54.1%
Female	45.9%
Ethnicity	
Asian/Pacific Islander	32.3%
African American	7.5%
Caucasian	55.8%
Hispanic	4.1%
Native or Alaskan American	0.3%
Income	
Less than $20,000	23.5%
$20,000 to $39,999	11.6%
$40,000 to $59,999	18.7%
$60,000 to $79,999	19.4%
$80,000 to $99,999	10.5%
$100,000 or above	16.3%
Mobile device always used for shopping	
Strongly agree (level 5)	15.6%
Agree (level 4)	18.7%
Neither (level 3)	16.0%
Disagree (level 2)	32.0%
Strongly disagree (level 1)	17.7%

Variables. We gathered H-Votes by asking the following question: *Do you find the above review helpful?* and prompting a YES or NO answer, just as (Site 1) does. The two latent foci of H-Votes (learning and decision-making) were measured with the question, *To what extent do you agree: the above review is helpful to me to LEARN (or MAKE A PURCHASE DECISION) about this product*, which prompted a response on a 5-point Likert scale (strongly disagree to strongly agree). For control variables, we used consumer age, ethnicity, income, and use of mobile phones for shopping. Table 4 shows the categories of age, ethnicity, and income, among others. We assessed consumers' use of mobile phones for shopping by asking the question, *I always use a mobile phone before or during shopping to get product/vendor/retailer information*, prompting a response on a 5-point Likert scale (strongly disagree to strongly agree).

5   Results

We summarize the results based on our three hypothesis pairs.

5.1   Influence of Helpfulness Voting Manipulation on Subsequent Votes (H1a and H1b)

To test the influence of helpfulness voting manipulation on subsequent votes across SEC categories or anchoring effect (H1), we used cross tabulations to account for the treatment conditions and consumer profile. The commonly used guidelines for cross tabulations [51] p. 626, [87] p. 734 are: (1) No more than 20% of the expected counts in cells are less than 5; and (2) All expected counts are 1 or greater. To minimize Type II errors (false negatives), we devised two-step crosstab analyses. First, we divided each consumer profile variable into upper and lower halves using the median (e.g., age, income), or into the majority and the non-majority (e.g., ethnicity). There are three treatments, so the number of crosstab cells was 2×3=6. Second, we conducted crosstab analyses with all possible pairs of the three treatments. By using 2×2 crosstabs, we maximized the number of statistically significant results that met the crosstab statistical requirements.

As shown in Table 5, anchoring appeared for at least one product in all the studied categories (Appendix B lists statistical details [1], [91] for Table 5). For positive reviews, anchoring effects appeared in 3 out of the 4 products (3 significant items for 3 goods in cells 1 & 2). For negative reviews, anchoring effects appeared for at least one product in each good type (6 significant items for two products in cells 5 & 6). Some consumers reversed their

Makoto Nakayama
Yun Wan

judgment of review IQ solely because they were presented with fake vote counts. These results support both H1a and H1b.

Table 5: Anchoring in the YES/NO evaluation of review helpfulness

Product	Positive Most Helpful Review	Negative Most Helpful Review
Search Good 1 Search Good 2	ethnicity** _H1a_ _cell 1_	_cell 4_
Experience Good 1 Experience Good 2	ethnicity* gender* _cell 2_	age**, gender**, ethnicity*** _H1b_ _cell 5_
Credence Good 1 Credence Good 2	_cell 3_	age*, gender*, mobile use* _cell 6_

*: α = .10, **: α = .05, ***: α = .01

5.2   Decision-Making and Learning are Different (H2a and H2b)

The variables for the degrees of helpfulness in learning and decision-making were statistically non-normal. Thus, we used Wilcoxon signed-ranked tests to examine their statistical differences (Table 6). These non-parametric ANOVA tests indicated that the two foci were different across all the SEC categories (i.e., each good in Table 6 has at least one significant z-value). This result supports H2a. H2b is supported by the trend in differentials, which increased from credence goods (one in four cases or 25% at α = .05) to search goods (three in four cases or 75% at α = .05), as shown in the rightmost column of Table 6.

Table 6: Differentials between learning and decision-making helpfulness for the same review

Product	Most Helpful Positive Review	Most Helpful Negative Review	Disparity
Search Good 1 Search Good 2	$z = -6.36$, $p = .000$ $z = -3.52$, $p = .000$	n.s $z = -3.09$, $p = .000$	75%
Experience Good 1 Experience Good 2	$z = -3.37$, $p = .001$ n.s	$z = -1.65$, $p = .098$ $z = -2.39$, $p = .017$	50%
Credence Good 1 Credence Good 2	n.s $z = -1.89$, $p = .058$	$z = -1.99$, $p = .047$ n.s	25%

The z- (the decision-making rating minus the learning rating) and p-values are the statistics for Wilcoxon signed-ranked test.

5.3   Influence of Helpfulness Voting Manipulation on Two Didden Dimensions (H3a and H3b)

Anchoring effects on learning and decision-making (H3) were tested with non-parametric one-way ANOVA tests, also known as Kruskal-Wallis tests. As shown in Table 7, there were significant anchoring effects in the two foci, learning and decision-making. (The details of these statistical results are in Appendix C.) Among the positive reviews for search goods and experience goods (cells 1 & 2 in Table 7), the decision-making (D) focus had 6 instances of anchoring, while the learning (L) focus had only 1 instance of anchoring. (Hereafter, we refer to instances of anchoring in decision-making and learning as D's and L's, respectively.) This result supports H3a (many more D's than L's). For the negative reviews of experience and credence goods (cells 5 & 6 in Table 7), we found 5 D's vs. 4 L's for experience goods and 0 D's and L's for credence goods. These results support H3b (similar number of D's and L's).

Table 7: Anchoring in learning and decision-making helpfulness

Product	Most Helpful Positive Review					Most Helpful Negative Review				
	age	ethnicity	gender	income	mobile use	Age	ethnicity	gender	income	mobile use
Search Good 1		L**	D**				L*, D*		D*	L*
Search Good 2			D**		cell 1			L***		cell 4
Experience Good 1	D**	D**		D**	D**			D*		
Experience Good 2					cell 2	L**, D**	L*, D**	L*, D**	D*	L*** cell 5
Credence Good 1			L**							
Credence Good 2	D**				cell 3					cell 6

L is learning helpfulness. D is decision-making helpfulness.
*: α = .10, **: α = .05, ***: α = .01

6   Discussions and Implications

Most importantly, we found the anchoring effect in consumers' evaluations of review IQ in (1) the YES/NO questions and (2) the grading scale for learning versus buying. The extent of anchoring depended on the review valence (positive or negative), SEC category (search, experience, or credence good), and consumer profile (e.g., age, gender, ethnicity). Anchoring can manifest instantly with purchase decisions or can surface over time through cumulative product learning. Table 8 summarizes the results of hypothesis testing.

Table 8: Summary of hypothesis testing

ID	What Was Tested	Result
H1	**Do fake summary IQ indicators affect YES/NO helpfulness votes?**	
H1a	They affect positive reviews for search and experience goods.	supported
H1b	They affect negative reviews for experience and credence goods.	supported
H2	**Do the two hidden foci exist?**	
H2a	They exist for all goods.	supported
H2b	Their differences increase in the order of search, experience, and credence goods.	supported
H3	**How do fake summary IQ indicators affect the two hidden foci differently?**	
H3a	They affect decision-making more than learning for positive reviews of search and experience goods.	supported
H3b	They have a similar effect for negative reviews of experience and credence goods.	supported

Practical consumer implications. Consumers are frequently not experts in purchasing high-tech electronics (search goods) and vitamin supplements (credence goods). This is partly because consumers purchase these goods relatively infrequently, partly because the goods often have complex attributes, and partly because judging a good can be challenging. As an example, some reviews claim that a printer from a given manufacturer suffers from paper jams. However, verifying this claim is not easy. Well-known online retailers offer many products in a given category to choose from; consumers typically see the products having mostly decent star ratings; and popular products have many reviews. Given these conditions, consumers often turn to the *most helpful* reviews prominently displayed on the first review page. This is perhaps the most common shopping experience among today's consumers.

As consumers read reviews and look for compelling reasons to buy a product, anchoring from H-Vote ratios-even though it is just one piece of information-can make a review look better or worse than it really is (H1a). Likewise, when consumers read reviews to find reasons against buying a product, anchoring can make the same review more or less convincing (H1b). Additionally, anchoring works differently in different demographics, including age, gender, and race (Table 6).

Consumers generally find a review beneficial either because it helps them learn more about the product or helps them decide to purchase it (H2a). For example, a consumer looking to buy a color laser printer or vitamin supplement would probably learn more about the product as they continue to research it. For goods like vitamins, further research may not help the consumer determine which product to buy; attempting to learn about these goods is likely futile. For products with tangible qualities, further research can make consumers more confident in their buying decisions. Learning about these goods is likely productive, and in this case a spread in review IQ is likely present between reviews which educate consumers and reviews that encourage them to buy the product (H2b). As a product's price increases, consumers become more likely to spend some time learning about the product, so *teaching* reviews are more important for expensive products than cheap products.

A given positive review with a fake high H-Vote ratio is more convincing than an identical review with a fake low H-Vote ratio, and this tendency is even stronger when the reviewed product has tangible or subjective benefits (H3a). The sensitivity to fake votes also depends on the consumer profile (Table 8), probably because each product has its own target audience. However, fake votes do not affect negative reviews of products like music albums and vitamins (H3b). In this case, someone's dislike does not equate to a potential buyer's dislike when it comes to subjective goods such as music albums and wine.

The good news is that fake helpfulness votes only fool certain consumers. Even so, consumers should be aware of fake votes because sellers can target consumers based on their profiles.

Detecting anchoring using the Guttman scale. The H-Votes are based on the binary mode of YES or NO, a classical Guttman scale [28]. Such a scale simplifies the response process, but it may oversimplify it [75]. For instance, a consumer leaning toward YES (or NO) who is subject to some anchoring may still give the same vote because the response scale hides the change in opinion. Thus, detecting anchoring effects is harder with Guttman scales than with Likert scales. Considering this possibility, the nature of support for H1a and H1b should be taken as the results obtained in the most challenging circumstances that have statistical significance. The anchoring changed consumer perception by only about one point when the reviews were evaluated based on objective criteria. When consumers read negative personal stories about experience and credence goods, though anchoring affected these goods more than search goods. That is, manipulations of H-Votes can more significantly affect negative reviews of experience

and credence goods than positive reviews of search goods. Most importantly, though, there is some evidence that anchoring affects review evaluations regardless of good type.

Meaning of the two hidden dimensions of helpfulness. Consumers generally perceive different values for learning and decision-making from the same review. This study used 12 *most helpful reviews* to test the hypotheses. If a review is extremely helpful or poorly written, the helpfulness for these two foci would most likely be similar because the scales of both foci would be impacted by their extremity. In this sense, we used the most challenging reviews to test the hypotheses. Yet, the two different foci are evident in the results. As hypothesized, products with tangible attributes produce more differences between learning and buying, while products with subjective attributes or unclear benefits produce fewer differences. The negative z values indicate the helpfulness for learning is higher than that for decision-making. This means that consumers are appropriately using more stringent criteria to judge the review helpfulness for deciding to buy a product than for learning about it. The learning focus has cumulative effects over time, while the decision-making focus has immediate effects on purchase behaviors. Thus, we argue that researchers should treat these foci differently. For some goods, one focus may be more important than the other. Complex goods, such as digital cameras and high-end electronic products, likely require more product learning unless the consumer is an expert on such goods. In contrast, frequently purchased goods, such as printer toner and popular dietary supplements, likely require a focus on decision-making. Product/marketing managers should incorporate the characteristics of the two foci in their product marketing strategies using reviews.

Consumer demographics and anchoring. A key message from Tables 5, 6, and 7 is that anchoring is specific to consumer demographics. For example, fake votes on a positive review for one music product tricked one ethnic group but not the other, but for a second music product, fake votes affected a third ethnic group. Some consumers can process more reviews than others in the same amount of time [42], and consumers with certain profiles find more value in some reviews over others. Anchoring in valuing reviews can manifest in present and future purchases specific to consumer demographics. For product marketers, this exploratory study demonstrates that is it possible to attract certain customer segments by having reviews favoring their products with falsely high H-Vote ratios. The customers with certain profiles are more susceptible to manipulated numbers.

As consumers, we should maintain a healthy skepticism of the evaluations we see for reviews. For example, certain types of music albums are *liked* by consumers with specific profiles. At the same time, these customers might be even more enthused with these albums after seeing artificially positive scores on reviews that strongly favor the albums. The positive biases of reviews are noted even in *USA Today*: "At one time on TripAdvisor, a travel review site, there were more than 50 million reviews and an average rating of 3.7 stars out of 5. Pretty high" [47]. Concerning travel reviews, one reporter commented, "People don't necessarily expect the truth when they click on a review site. Truthy is good enough" [21]. Consumers have good reason to be aware that anchoring can cloud their purchase judgments.

Tradeoffs in judging reviews with YES-or-NO scales. As researchers, we should be mindful that review evaluations are not objective. Just like public opinions aroused or swayed by a few vocal groups and entities, a review's helpfulness could be manipulated just the same. In particular, the Guttman scale commonly used in rating the helpfulness of reviews should raise interesting debates. As Masoff [50] discussed extensively, the Likert and Guttman scales have different assumptions and aims. The two latent foci of review helpfulness follow non-normal distributions. The star rating of products is known to follow a J-distribution [34]. Thus, not using an interval scale may have been an effective decision for (Site 1). At the same time, our study revealed the *hidden* values of reviews beyond the YES/NO basis. Overall, there seem to be theoretical and practical tradeoffs when choosing a measurement scale for review helpfulness.

7   Managerial Implications

Identifying fake reviews and other review manipulation is a constant battle for online vendors with both economical and legal implications. If they cannot effectively control review manipulation, consumers will eventually lose trust in reviews. In certain circumstances, tolerance of such behavior may be seen as collaboration and lead to class lawsuits. Either situation would damage a company's reputation and profits.

Our study shed some light on how manipulation affects online reviews. First, we confirmed that anchoring affects the process of rating online reviews for helpfulness, a warning to anyone who plans to use such an IQ rating as a non-biased input in other review data-mining.

Second, we revealed that different product categories are susceptible to different manipulations of the helpfulness rating. Thus, online vendors should spend time designing measures to counteract or mitigate the influence of anchoring in specific product categories.

8 Limitations and Future Research

This study has a few limitations. The survey questionnaire was administered in 2012, however we have not seen significant changes in how reviews are rated at most websites. We did not consider consumers' actual purchase histories and product expertise. We used ordinary, common products with which many consumers likely have at least some purchase experience and knowledge. Future studies may want to explore the relationship between actual purchase records and the anchoring effect. A neurocognitive or neuromarketing, [12], [25] approach could advance beyond the findings of study.

Reviews of subjective products, such as music albums, were affected by fake votes differently depending on the consumer's age, gender, and ethnicity. This result is intuitive because we know different consumer groups prefer certain types of music. Using larger samples, we could refine the findings of this study. Future studies should also look into the sensitivity analysis between anchoring, product interests, and consumer profile.

This exploratory study used 12 reviews of bestselling products as benchmarks, but they may not be sufficiently representative. To increase the variety of reviews, future studies could sample more products in each SEC good category to ensure the results are statistically sound.

Another rewarding direction would be to investigate the relationship between the anchoring effect and the textual characteristics of reviews. For example, future studies could classify reviews by their readability and their proportions of objective/subjective content, then evaluate the influence of anchoring. Other possible research topics include studying how anchoring affects repurchase intention and perception of product quality.

9 Conclusion

Popular products in online stores often have overwhelming numbers of reviews. One simple way to aid consumers in analyzing reviews is to allow them to vote YES or NO on whether a review is helpful, collecting the votes to signal the quality of the review. However, online reviews are often manipulated, so how do these fake votes affect consumers? To clarify this, we asked two research questions: First, to what extent do fake votes anchor consumers' judgments of review IQ? Second, do fake votes affect both review judgments and the process of researching product and deciding to purchase it?

This exploratory study is one of the first to report that some consumers, anchored by fake vote counts, reversed their Yes/No vote on review IQ. While previous studies assumed that vote counts are reflections of perceived review IQs, our results show that this assumption is not always valid. Not only did fake votes reverse the Yes/No decision of some consumers, but they also affected the process of review judgment and product research. Specifically, when votes were manipulated in a positive review of a search or experience good, this affects the rational or logical reasons to buy the product. When votes are manipulated in a negative review on an experience or credence good, this changes a subjective or emotional opinion as to why consumers should not buy the product. These influences depend on consumer profiles because the nature of goods vary and because specific attributes of products are different. Thus, anchoring of fake review IQ votes can influence not only subsequent votes but also the consumers' product research and purchase decision-making.

Our findings provide important insights for future analysis of user-generated content and their evaluations. Previous studies tended to use public-polled data, such as helpfulness rating, as a benchmark or objective inputs. Our study found that anchoring affects these data and discussed how anchoring affects various product categories and consumers' processes of product learning and purchase decision-making. Consumers, retailers, and vendors should take a fresh look at review helpfulness vote counts beyond something trustworthy.

Websites List

Site 1: Amazon.com, Inc
http://www.amazon.com

Site 2: Dealsea.com
http://www.dealsea.com

Site 3: Whitepages.com
http://www.whitepages.com

References

[1] N. Ahituv, S. Neumann and M. Zviran. Factors affecting the policy for distributing computing resources. MIS Quarterly, vol. 13, no. 4, pp. 389-401, 1989.

[2] I. Ajzen, The theory of planned behavior. Organizational Behavior and Human Decision Processes, vol. 50, no. 2, pp. 179-211, 1991.

[3] E. Anderson and D. Simester, Mind your pricing cues, Harvard Business Review, vol. 81, no. 9, pp. 96-103, 2003.

[4] S. E. Ante. (June 3, 2016) Amazon: Turning consumer opinions into gold. Bloomberg, L.P. [Online]. Available: http://www.bloomberg.com/news/articles/2009-10-15/amazon-turning-consumer-opinions-into-gold.

[5] H. Baek, J. Ahn and Y. Choi, Helpfulness of online consumer reviews: Readers' objectives and review cues, International Journal of Electronic Commerce, vol. 17, no. 2, pp. 99-126, 2012.

[6] J. Bateman, I. H. Langford, A. P. Jones, and G. N. Kerr, Bound and path effects in double and triple bounded dichotomous choice contingent valuation, Resource and Energy Economics, vol. 23, no. 3, pp. 191-213, 2001.

[7] N. N. Bechwati and L. Xia. Do computers sweat? The impact of perceived effort of online decision aids on consumers' satisfaction with the decision process, Journal of Consumer Psychology, vol. 13, no. 1-2, pp. 139-148, 2003.

[8] T. Betsch and S. Haberstroh, The Routines of Decision Making. Mahwah, NJ: Lawrence Erlbaum, 2004.

[9] T. Betsch, S. Haberstroh and C. Hohle, Explaining routinized decision making a review of theories and models, Theory & Psychology, vol. 12, no. 4, pp. 453-488, 2002.

[10] J.R. Bettman, E.J. Johnson and J.W. Payne, Consumer decision making, in Handbook of Consumer Behavior (T.S. Robertson and H.H. Kassarjian, Eds.). Englewood Cliffs, NJ: Prentice Hall, 1991, pp. 50-84.

[11] K. L. Blankenship, D. T. Wegener, R.E. Petty, B. Detweiler-Bedell, and C. L. Macy, Elaboration and consequences of anchored estimates: An attitudinal perspective on numerical anchoring, Journal of Experimental Social Psychology, vol. 44, no. 6, pp. 1465-1476, 2008.

[12] D. Booth and R. P. Freeman, Mind-reading versus neuromarketing: How does a product make an impact on the consumer? Journal of Consumer Marketing, vol. 31, no. 3, pp. 177-189, 2014.

[13] R. J. Brodie, A. Ilic, B. Juric, and L. Hollebeek, Consumer engagement in a virtual brand community: An exploratory analysis, Journal of Business Research, vol. 66, no. 1, pp. 105-114, 2013.

[14] Q. Cao, W. Duan and Q. Gan, Exploring determinants of voting for the helpfulness of online user reviews: A text mining approach, Decision Support Systems, vol. 50, no. 2, pp. 511-521, 2011.

[15] M. R. Darby and E. Karni, Free competition and the optimal amount of fraud, Journal of Law and Economics, vol. 16, no. 1, pp. 67-88, 1973.

[16] W. K. Darley, C. Blankson and D. J. Luethge, Toward an integrated framework for online consumer behavior and decision making process: A review, Psychology & Marketing, vol. 27, no. 2, pp. 94-116, 2010.

[17] W. K. Darley and R. E. Smith, Advertising claim objectivity: Antecedents and effects, Journal of Marketing, vol. 57, no. 4, pp. 100-113, 1993.

[18] F. D. Davis, Perceived usefulness, perceived ease of use, and user acceptance, MIS Quarterly, vol. 13, no. 3, pp. 319-340, 1989.

[19] A. Donner and M. Eliasziw, Statistical implications of the choice between a dichotomous or continuous trait in studies of interobserver agreement, Biometrics, vol. 50, no. 2, pp. 550-555, 1994.

[20] H. J. Einhorn, The use of nonlinear, noncompensatory models in decision making, Psychological Bulletin, vol. 73, no. 3, pp. 221-230, 1970.

[21] C. Elliott. (2013, September) Do fake travel reviews really matter?. Elliot. [Online]. Available: http://elliott.org/elliotts-email/do-fake-travel-reviews-really-matter/

[22] J. Engel, R. Blackwell and P. Miniard, Consumer Behavior, 5th ed. Hinsdale, IL: Dryden, 1986.

[23] J. F. Engel, D. T. Kollatt and R. D. Blackwell, Consumer Behavior, 3rd ed. Hinsdale, IL: Dryden Press, 1978.

[24] N. Epley and T. Gilovich, The anchoring-and-adjustment heuristic Why the adjustments are insufficient, Psychological Science, vol. 17, no. 4, pp. 311-318, 2006.

[25] Z. Eser, F. B. Isin and M. Tolon, Perceptions of marketing academics, neurologists, and marketing professionals about neuromarketing, Journal of Marketing Management, vol. 27, no. 7-8, pp. 854-868, 2011.

[26] M. Fishbein and I. Ajzen, Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Reading. MA: Addison-Wesley, 1975.

[27] S. Frederick, D. Kahneman and D. Mochon, Elaborating a simpler theory of anchoring, Journal of Consumer Psychology, vol. 20, no. 1, pp. 17-19, 2010.

[28] L. Guttman, A basis for scaling qualitative data, American Sociological Review, vol. 9, no. 2, 1944, pp. 139-150.

[29] G. Haubl and V. Trifts, Consumer decision making in online shopping environments: The effects of interactive decision aids, Marketing Science, vol. 19, no. 1, pp. 4-21, 2000.

[30] H. D. Ho, S. Ganesan and H. Oppewal, The impact of store-price signals on consumer search and store evaluation, Journal of Retailing, vol. 87, no. 2, pp. 127-141, 2011.

[31] Y. C. Hsieh, H. C. Chiu and M. Y. Chiang, Maintaining a committed online customer: A study across search-experience-credence products, Journal of Retailing, vol. 81, no. 1, pp. 75-82, 2005.

[32] N. Hu, I. Bose, Y. Gao, and L. Liu, Manipulation in digital word-of-mouth: A reality check for book reviews, Decision Support Systems, vol. 50, no. 3, pp. 627-635, 2011.

[33] N. Hu, I. Bose, N.S. Koh, and L. Liu, Manipulation of online reviews: An analysis of ratings, readability, and sentiments, Decision Support Systems, vol. 52, no. 3, pp. 674-684, 2012.

[34] N. Hu, J. Zhang and P.A. Pavlou, Overcoming the J-shaped distribution of product reviews, Communications of the ACM, vol. 52, no. 10, pp. 144-147, 2009.

[35] P. Huang, N.H. Lurie and S. Mitra. Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods, Journal of Marketing, vol. 73, no. 2, pp. 55-69, 2009.

[36] W.J. Hutchinson and E.M. Eisenstein, Consumer learning and expertise, in Handbook of Consumer Psychology (C.P. Haugtvedt, P.M. Herr and F.R. Kardes, Eds.). New York, NY: Psychology Press, Taylor & Francis Group, 2008, pp. 103-132.

[37] S. L. Jarvenpaa, The effect of task demands and graphical format on information processing strategies, Management Science, vol. 35, no. 3, pp. 285-303, 1989.

[38] N. Jindal and B. Liu, Opinion spam and analysis, in Proceedings 2008 International Conference on Web Search and Data Mining, Stanford, CA: ACM, 2008, pp. 219-230.

[39] D. J. Kim, D. L. Ferrin and H. R. Rao, A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents, Decision Support Systems, vol. 44, no. 2, 2008, pp. 544-564.

[40] R. Kohli, S. Devaraj and A. Mahmood, Understanding determinants of online consumer satisfaction: A decision process perspective, Journal of Management Information Systems, vol. 21, no. 1, pp. 115-135, 2004.

[41] N. Korfiatis, E. García-Bariocanal and S. Sánchez-Alonso, Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content, Electronic Commerce Research and Applications, vol. 11, no. 3, pp. 205-217, 2012.

[42] S. Krishen, R. L. Raschke and P. Kachroo, A feedback control approach to maintain consumer information load in online shopping environments, Information & management, vol. 48, no. 8, pp. 344-352, 2011.

[43] W. Kruglanski and T. Freund, The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring, Journal of Experimental Social Psychology, vol. 19, no. 5, pp. 448-468, 1983.

[44] K. K. Y. Kuan, P. Prasarnphanich, K. L. Hui, and H. Y. Lai, what makes a review voted? An empirical investigation of review voting in online review systems, Journal of the Association for Information Systems, vol. 16, no. 1, pp. 48-71, 2015.

[45] R.Y. Lau, S. Liao, R.C.W. Kwok, K. Xu, Y. Xia, and Y. Li, Text mining and probabilistic language modeling for online review spam detecting, ACM Transactions on Management Information Systems, vol. 2, no. 4, pp. 1-30, 2011.

[46] R. Y. K. Lau, S. Y. Liao, R. C. W. Kwok, K. Xu, Y. Xia, and Y. Li, Text mining and probabilistic language modeling for online review spam detection, ACM Transactions on Management Information Systems, vol. 2, no. 4, pp. 1-30, 2012.

[47] R. Lewis. (2013, February) Money quick tips: Can you trust online reviews?. USA Today. [Online]. Available: http://www.usatoday.com/story/money/personalfinance/2013/02/12/money-quick-tips-online-reviews/1911487/

[48] E. P. Lim, V. A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, Detecting product review spammers using rating behaviors, in Proceedings 19th ACM International Conference on Information and Knowledge Management, Toronto, Canada, 2010, pp. 939-948.

[49] G. Linden, B. Smith and J. York, Amazon.com recommendations: Item-to-item collaborative filtering, IEEE Internet Computing, vol. 7, no. 1, pp. 76-80, 2003.

[50] R. W. Massof, Likert and Guttman scaling of visual function rating scale questionnaires, Ophthalmic Epidemiology, vol. 11, no. 5, pp. 381-399, 2004.

[51] D. S. Moore and G. P. McCabe, Introduction to the Practice of Statistics. New York: W. H. Freeman and Company, 2003.

[52] S. Mudambi and D. Schuff, What makes a helpful online review? A study of customer reviews on amazon.com, MIS Quarterly, vol. 34, no. 1, pp. 185-200, 2010.

[53] A. Mukherjee, B. Liu and N. Glance, Spotting fake reviewer groups in consumer reviews, in Proceedings 21st International Conference on World Wide Web, Lyon, France, pp. 191-200, 2012.

[54] M. Nakayama, Y. Wan and N. Sutcliffe, How dependent are consumers on others when making their shopping decisions? Journal of Electronic Commerce in Organizations, vol. 9, no. 4, pp. 1-21, 2011.

[55] P. Nelson, Information and consumer behavior, Journal of Political Economy, vol. 78, no. 2, pp. 311-329, 1970.

[56] P. Nelson, Advertising as information, The Journal of Political Economy, vol. 82, no. 4, pp. 729-754, 1974.

[57] B.R. Newell and A. Bröder, Cognitive processes, models and metaphors in decision research, Judgment and Decision Making, vol. 3, no. 3, pp. 195-204, 2008.

[58] O. Oeusoonthornwattana and D.R. Shanks, I like what I know: Is recognition a non-compensatory determiner of consumer choice, Judgment and Decision Making, vol. 5, no. 4, pp. 310-325, 2010.

[59] T. Pachur, A. Bröder and J.N. Marewski, The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? Journal of Behavioral Decision Making, vol. 21, no. 2, pp. 183-210, 2008.

[60] Y. Pan and J.Q. Zhang, Born unequal: A study of the helpfulness of user-generated product reviews, Journal of Retailing, vol. 87, no. 4, pp. 598-612, 2011.

[61] D. H. Park and S. Kim, The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews, Electronic Commerce Research and Applications, vol. 7, no. 4, pp. 399-410, 2008.

[62] R. E. Petty and J. T. Cacioppo, The elaboration likelihood model of persuasion, Advances in experimental social psychology, vol. 19, no. 1, pp. 123-205, 1986.

[63] R.E. Petty and D.T. Wegener, The elaboration likelihood model: Current status and controversies, in Dual-process Theories in Social Psychology (S. Chaiken and Y. Trope, Eds.). New York: Guilford Press, 1999, pp. 41-72.

[64] PowerReviews. (2011, June) The 2011 social shopping study. Power Reviews. [Online]. Available: http://www.powerreviews.com/assets/download/Social_Shopping_2011_Brief1.pdf

[65] S. Prawesh and B. Padmanabhan, The most popular news recommender: Count amplification and manipulation resistance, Information Systems Research, vol. 25, no. 3, pp. 569-589, 2014.

[66] Reevoo. (2013, March) The reevoo consumer purchasing report. Reevoo. [Online]. Available: http://www.reevoo.com/resources/research/reevoo-consumer-purchasing-report-september-2011

[67] M. Reitsma-van Rooijen and D. D. L. Daamen, Subliminal anchoring: The effects of subliminally presented numbers on probability estimates, Journal of Experimental Social Psychology, vol. 42, no. 3, pp. 380-387, 2006.

[68] S. Rezaei, Segmenting consumer decision-making styles (CDMS) toward marketing practice: A partial least squares (PLS) path modeling approach, Journal of Retailing and Consumer Services, vol. 22, pp. 1-15, 2015.

[69] S. Shin, S. Misra and D. Horsky, Disentangling preferences and learning in brand choice models, Marketing Science, vol. 31, no. 1, pp. 115-137, 2012.

[70] H. Simon, The New Science of Management Decision. Englewood Cliffs, NJ: Prentice Hall, 1977.

[71] H. A. Simon, A behavioral model of rational choice, The Quarterly Journal of Economics, vol. 69, no. 1, pp. 99-118, 1955.

[72] H. A. Simon, Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization. New York, NY: Macmillan, 1957.

[73] E. Sirakaya and A.G. Woodside, Building and testing theories of decision making by travelers, Tourism Management, vol. 26, no. 6, pp. 815-832, 2005.

[74] C. Smallman and K. Moore, Process studies of tourists' decision-making, Annals of Tourism Research, vol. 37, no. 2, pp. 397-422, 2010.

[75] J. S. Smith, The use of the Guttman scale in market research, The Incorporated Statistician, vol. 10, no. 1, pp. 15-28, 1960.

[76] E. K. Sproles and G. B. Sproles. Consumer Decision-Making Styles as a Function of Individual Learning Styles. Journal of Consumer Affairs, vol. 24, no. 1, pp. 134-147, 1990.

[77] J. Sussin and E. Thompson. (2013, January) Solve the problem of fake online reviews or lose credibility with consumers. Gartner. [Online]. Available: https://www.gartner.com/doc/2313315/solve-problem-fake-online-reviews

[78] J. Swait, A non-compensatory choice model incorporating attribute cutoffs, Transportation Research Part B: Methodological, vol. 35, no. 10, pp. 903-928, 2001.

[79] A. Tversky, Elimination by aspects: A theory of choice, Psychological Review, vol. 79, no. 4, pp. 281-299, 1972.

[80] A. Tversky and D. Kahneman, Judgment under uncertainty: Heuristics and biases, Science, vol. 185, no. 4157, pp. 1124-1131, 1974.

[81] B. Wansink, R.J. Kent and S.J. Hoch, An anchoring and adjustment model of purchase quantity decisions, Journal of Marketing Research, vol. 35, no. 1, pp. 71-81, 1998.

[82] T. Wegener, R. E. Petty, K.L. Blankenship, and B. Detweiler-Bedell, Elaboration and numerical anchoring: Breadth, depth, and the role of (non-)thoughtful processes in anchoring theories, Journal of Consumer Psychology, vol. 20, no. 1, pp. 28-32, 2010.

[83] T. Wegener, R. E. Petty, K.L. Blankenship, and B. Detweiler-Bedell, Elaboration and numerical anchoring: Implications of attitude theories for consumer judgment and decision making, Journal of Consumer Psychology, vol. 20, no. 1, pp. 5-16, 2010.

[84] T. Wegener, R. E. Petty, B.T. Detweiler-Bedell, and W. B. G. Jarvis, Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness, Journal of Experimental Social Psychology, vol. 37, no. 1, pp. 62-69, 2001.

[85] M. P. Welsh and G. L. Poe, Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach, Journal of Environmental Economics and Management, vol. 36, no. 2, pp. 170-185, 1998.

[86] A. Wright and J. G. Lynch, Jr., Communication effects of advertising versus direct experience when both search and experience attributes are present, The Journal of Consumer Research, vol. 21, no. 4, pp. 708-718, 1995.

[87] D. Yates, D. Moore and D.S. Starnes, The Practice of Statistics. New York: Macmillan, 1999.

[88] T. Zhang and D. Zhang, Agent-based simulation of consumer purchase decision-making and the decoy effect, Journal of Business Research, vol. 60, no. 8, pp. 912-922, 2007.

[89] Y. Zhao, S. Yang, V. Narayan, and Y. Zhao, Modeling consumer learning from online product reviews, Marketing Science, vol. 32, no. 1, pp. 153-169, 2013.

[90] Y. Zhao, Y. Zhao and K. Helsen, Consumer learning in a turbulent market environment: Modeling consumer choice dynamics after a product-harm crisis, Journal of Marketing Research (JMR), vol. 48, no. 2, pp. 255-267, 2011.

[91] M. Zviran and W.J. Haga, Password security: An empirical study, Journal of Management Information Systems, vol. 15, no. 4, pp. 161-185, 1999.

Appendix A: Survey Instruments

Here is [Product Name] from Amazon. Please go to the next page and let us know if you feel the review about this printer is helpful.

<The product image is displayed just as in Amazon.com right below the above message.>

Amazon Review – [Product Name]

<Just as shown at Amazon.com, a review is shown with original or manipulated YES vote counts.>

Do you find above review helpful?

◯ Yes

◯ No

To what extent do you agree: Above review is helpful to me to LEARN about this product.

◯ Strongly disagree ◯ Disagree ◯ Neutral ◯ Agree ◯ Strongly agree

To what extent do you agree: Above review is helpful to me to MAKE A PURCHASE DECISION about this product.

◯ Strongly disagree ◯ Disagree ◯ Neutral ◯ Agree ◯ Strongly agree

Appendix B: Statistical Details for Table 5

Product	Valence	Profile	Chi-Square	Nominal/Ordinal Statistics**
Search Good 1	Positive	Caucasian (all*)	χ = 6.29, *df* = 2, p = 0.045	CV = 0.270, τ = 0.073, UC = 0.083
Search Good 1	Positive	Caucasian (LN)	χ = 6.32, *df* = 1, p = 0.018	CV = 0.270, τ = 0.073, UC = 0.083
Experience Good 1	Positive	Caucasian (all)	χ = 5.37, *df* = 2, p = 0.076	CV = 0.254, τ = 0.065, UC = 0.082
Experience Good 2	Positive	Female (HL)	χ = 4.06, *df* = 1, p = 0.067	CV = 0.307, τ = 0.094, UC = 0.070
Experience Good 2	Negative	Younger (all)	χ = 9.96, *df* = 2, p = 0.017	τc = 0.227, SD = 0.171
Experience Good 2	Negative	Non-Caucasian (all)	χ = 7.81, *df* = 2, p = 0.022	CV = 0.341, τ = 0.116, UC = 0.102
Experience Good 2	Negative	Younger (HL)	χ = 6.16, *df* = 1, p = 0.021	τb = 0.299, SD = 0.295
Experience Good 2	Negative	Female (LN)	χ = 5.99, *df* = 1, p = 0.022	CV = 0.339, τ = 0.115, UC = 0.102
Experience Good 2	Negative	Non-Caucasian (HL)	χ = 7.94, *df* = 1, p = 0.007	CV = 0.432, τ = 0.182, UC = 0.163
Credence Good 1	Negative	Younger (HL)	χ = 3.63, *df* = 1, p = 0.070	τb = -0.206, SD = -0.195
Credence Good 1	Negative	Mobile, Lower (HL)	χ = 3.52, *df* = 1, p = 0.100	τb = -0.236, SD = 0.045
Credence Good 1	Negative	Male (LN)	χ = 3.42, *df* = 1, p = 0.094	CV = 0.241, τ = 0.058, UC = 0.046

*: Treatments: all: high-low-none   HL: high-low   HN: high-none   LN: low-none
**: Legends: CV: Cramer's V   τ: Goodman and Kruskal's tau   UC: uncertainty coefficient
SD:Summer's   τb: Kendall's tau-b τc: Kendall's tau-c   p: two-sided p value, exact results for 2x2 crosstab, Monte Carlo simulation with 10,000 samples for all treatments

Makoto Nakayama
Yun Wan

Appendix C: Statistical Details for Table 7

Product	Most Helpful Positive Review	Most Helpful Negative Review
Search Good 1	Learning: ethnicity (Caucasian), χ=6.52, df = 2, p = 0.035	Learning: ethnicity (Non-Caucasian), χ=5.16, df = 2, p = 0.075
	Decision: gender (male), χ=5.74, df = 2, p = 0.035	Learning: mobile use (High), χ=5.03, df = 2, p = 0.075
		Decision: ethnicity (Non-Caucasian), χ=5.16, df = 2, p = 0.072
		Decision: income (High), c=5.44, df = 2, p = 0.066
Search Good 2	Decision: gender (male), χ=6.77, df = 2, p = 0.036	Learning: gender (female), χ=9.74, df = 2, p = 0.008
Experience Good 1	Decision: age (young), χ=6.59, df = 2, p = 0.036	Decision: gender (male), χ=5.35, df = 2, p = 0.072
	Decision: ethnicity (Caucasian), χ=8.54, df = 2, p = 0.012	
	Decision: income (High), c=6.43, df = 2, p = 0.037	
	Decision: mobile use (High), χ=8.31, df = 2, p = 0.014	
Experience Good 2		Learning: age (young), χ=7.24, df = 2, p = 0.025
		Learning: ethnicity (Non-Caucasian), χ=5.33, df = 2, p = 0.071
		Learning: gender (female), χ=5.19, df = 2, p = 0.074
		Learning: mobile use (Low), χ=9.70, df = 2, p = 0.007
		Decision: age (young), χ=6.69, df = 2, p = 0.035
		Decision: ethnicity (Non-Caucasian), χ=8.05, df = 2, p = 0.015
		Decision: gender (female), χ=8.67, df = 2, p = 0.013
		Decision: income (High), c=5.16, df = 2, p = 0.072
		Decision: mobile use (Low), χ=9.70, df = 2, p = 0.007
Credence Good 1	Learning: gender (male), χ=7.54, df = 2, p = 0.020	
Credence Good 2	Decision: age (young), χ=4.80, df = 2, p = 0.082	

P-value: Monte Carlo simulation with the number of 10,000 samples

Makoto Nakayama
Yun Wan