Uncovering the Coexistence of Assimilation and Contrast Effects in Hedonic Sequences

Tanuka Ghoshal
*Indian School of Business*

Eric Yorkston
*Texas Christian University*

Joseph C. Nunes
*University of Southern California, Los Angeles*

Peter Boatwright
*Carnegie Mellon University, pbhb@andrew.cmu.edu*

Follow this and additional works at: [http://repository.cmu.edu/tepper](http://repository.cmu.edu/tepper)
Uncovering the Coexistence of Assimilation and Contrast Effects in Hedonic Sequences

TANUKA GHOSHAL
ERIC YORKSTON
JOSEPH C. NUNES
PETER BOATWRIGHT

Tanuka Ghoshal is Assistant Professor of Marketing at the Indian School of Business, Gachibowli, Hyderabad 500032, India. Email: Tanuka_Ghoshal@isb.edu

Eric Yorkston is Associate Professor of Marketing at the Neeley School of Business, Texas Christian University, TCU Box 298530, Fort Worth, TX 76129. Email: e.yorkston@tcu.edu

Joseph C. Nunes is Associate Professor of Marketing at the Marshall School of Business, University of Southern California, Los Angeles, CA 90089-0443. Email: JNunes@marshall.usc.edu

Peter Boatwright is Associate Professor of Marketing at the Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213. Email: boatwright@cmu.edu

Questions should be directed to Tanuka Ghoshal at Tanuka_Ghoshal@isb.edu.
ABSTRACT

Most judgments consumers make are parts of sequences and hence unlikely to be free of context effects. Assimilation (contrast) refers to a positive (negative) relationship between the value people place on the context and the value they place on the target stimulus. A general presupposition for much of the work on assimilation and contrast is that one or the other, determined by various factors, occurs. We propose that assimilation and contrast can co-occur within a sequence of experiences and present a hierarchical Bayesian model separating these effects within a unique real-world data set. We find that while assimilation effects influence overall sequence means, contrast effects are simultaneously evident between adjacent items and after extremes within a sequence. This work is the first empirical demonstration of hedonic contrast using real-world data, and the only work thus far to identify and separate assimilation and contrast effects within the same sequence of evaluations.

Keywords: Context effects, assimilation, contrast, hedonic sequence, hierarchical Bayes
According to wine enthusiasts, a wine's quality assessment is more accurate when performed alongside several other wines, in what are known as tasting “flights.” Wines may be deliberately selected for their vintage (“horizontal” tasting) or proceed from a single winery (“vertical” tasting), but most often both the vintage and winery are masked to promote a purportedly “unbiased” analysis. But how unbiased is it to taste several wines in sequence? Wine, beer, whiskey, and other spirits are often tasted and judged in flights. Most judgments that consumers make everyday are also part of sequences. Whether test driving cars, evaluating courses of a menu, dating (in particular, speed-dating), or completing a customer satisfaction survey, it is difficult to imagine how the results from one evaluation would not affect the next.

It is widely known that judgments are not context-free (Dato-on and Dahlstrom 2003; Stapel and Winkielman 1998). The literature in marketing is replete with examples of how the context in which a stimulus is embedded can have a significant impact on people’s judgment of that stimulus. The two context effects that have been most reliably demonstrated in psychology and marketing are assimilation and contrast. Assimilation refers to a positive relationship between the value people place on the contextual stimuli surrounding a target and the value they place on that target itself. Contrast refers to a negative relationship between these two values (Martin, Seta, and Crelia 1990; Sherman et al. 1978).

Many factors have been shown to be responsible for determining whether assimilation or contrast effects occur. For example, if the evaluation of the target is at more of an automatic, spontaneous, or holistic level, assimilation is assumed to result, while if the evaluation is in a more deliberative mode where cognitive resources are both available and expendable, contrast effects are more likely (Martin 1986; Myers-Levy and Tybout 1997; Bickart 1993). The general conclusion is that assimilation is the “default” outcome in contextual influence and that the
natural tendency is for people to assimilate while contrast is the outcome of a more effortful process in high-involvement situations.

Other work has focused on the specificity and typicality of the target stimuli as determinants of whether assimilation or contrast occurs. When the target is ambiguous, then the context information is more likely to be used as an interpretative frame and the result is assimilation (Stapel and Koomen 1997; Stapel and Winkielman 1998). Research examining the relationship between the target and the contextual information has found that if items are perceived to be similar (domain match), contrast effects are enhanced, while if they are perceived to be dissimilar (domain mismatch), assimilation effects are more likely to occur (Sherif and Hovland 1961). Schwarz and Bless (1992) similarly argue that the way in which a target stimulus is categorized will determine whether assimilation or contrast occurs. They argue, in general, that if a target is included in a mental representation of the category, assimilation effects are more likely to occur. Conversely, if the target is excluded, contrast effects are more likely. Their work argues that either assimilation or contrast effects will manifest themselves, as stimuli cannot be both similar and dissimilar or simultaneously exist both within and outside of a category.

A more recent stream of research focuses on characterizing context effects in sequences of hedonic experiences (i.e., incidents of pleasure or pain). The results have been mixed. For example, Raghunathan and Irwin (2001) document contrast effects in respondents’ evaluations of descriptions of vacation spots and cars when there is a domain match but assimilation effects in cases of a domain mismatch, consistent with the early work by Sherif and Hovland (1961). Yet Novemsky and Ratner (2003) find that although individuals expect and predict contrast effects, evaluations provided at the time of the actual experience provide no evidence of contrast effects. Their work conflicts with research by Zellner and colleagues who find contrast effects not only
in the retrospective evaluation of gourmet versus ordinary coffee and specialty versus regular beer (Zellner, Kern, and Parker 2002), but in actual tasting experiences of fruit juices of varying concentrations (Zellner et al. 2003). Our work sheds light on why some researchers might observe hedonic contrast while others do not, which we explain below.

What has been consistent with respect to the previous research on assimilation and contrast across sequences of experiences is that it is always assumed that only one or the other takes place and that characteristics of the context, such as domain match (Meyers-Levy and Sternthal 1993, Raghunathan and Irwin 2001), product knowledge (Bickart 1993), or context set range (Lynch, Chakravarti, and Mitra 1991) dictate which one occurs. We propose that within a sequence of evaluations, specifically comparisons, both assimilation and contrast can co-occur and that previous analyses focusing on, and thus testing for one exclusively, may have failed to pick up the other.

In this research, we demonstrate how assimilation and contrast can co-occur in the same sequence by separating these effects with the aid of a hierarchical Bayesian model. We find that while assimilation effects influence overall group or sequence means, contrast effects are also evident among adjacent items and after extremes within a sequence. Therefore, contrary to Novemsky and Ratner (2003), we are able to document contrast effects in real-time sequential hedonic evaluations. Additionally, by documenting how assimilation and contrast effects can co-occur, and how controlling for one reveals the other, our work contributes to numerous research streams including work on sequential evaluation, taste, and the evaluation of hedonic experiences.

The remainder of this paper is organized as follows. First, we describe the rich, real-world data set that allows us to simultaneously test for assimilation and contrast. Next, we
present our hypotheses, which are based on the relevant research in sequential evaluation, taste, and context effects involving hedonic experiences. We then present the preliminary data analysis that guided us in the formulation of the model before proceeding to our discussion of the formal model. We present the results and conclude by discussing managerial implications and opportunities for future research.

DATA DESCRIPTION

The objective of this paper is to study context effects in a real-world application of sequential hedonic evaluation. To this end, we obtained eight years of judging data from the Bluebonnet Brew-off, a national beer brewing competition held annually in the Dallas/Fort Worth Metroplex. From 2000 to 2007, more than 900 brewers entered a total of 5,060 beers in 23 style categories (consisting of 107 subcategories). In order to judge such a large number of beers and yet avoid judges experiencing taste or “palate fatigue,” beers were grouped into “flights” ranging from five to 13 beers. There are 688 flights in our data. Within a flight, all beers belonged to the same (style) category, but many flights contained beers from more than one subcategory. For example, Brown Porter, Robust Porter, and Baltic Porter are all subcategories of the Porter (style) category and could appear within the same flight. All the beers within a flight that belong to the same subcategory form a “sub-flight,” and beers from the same category are randomly assigned to flights. Table 1 contains summary statistics describing the flights and sub-flights in our dataset.

Insert table 1 about here
Each flight was rated by two (occasionally, three) independent judges who sampled each of the beers within the flight in the same order, progressing from lighter to darker subcategories in order to preserve palate integrity and allow judges to fairly evaluate each beer. Judges in our sample have one of four levels of certification, which consist of (in ascending order): Apprentice, Recognized, Certified, and National, depending on the level of the Beer Judge Certification Program (BJCP) they have completed. In the current BJCP scoring system, each beer is evaluated on a 50-point scale comprised of 20 points for flavor, 12 points for aroma, three points for appearance, five points for mouthfeel, and 10 points for overall impression, which is an evaluation of the overall drinking pleasure of the entry (Appendix). Note that the data recorded and retained by the organization, the same that has been provided to us, contain the summary score out of 50 for each beer rather than the points on each individual component. Judges are advised to rinse their mouths and to cleanse their palates with bread or salt-free crackers between each evaluation of consecutive beers. Each score is recorded while tasting the specific beer and written down prior to moving on to the next beer within the flight.

The concurrent scoring of each beer during tasting ensures that the evaluations in our sample are actual real-time experience data and not recalled experience or retrospective evaluation data. After all beers within a flight are judged, each beer is assigned a rank within the flight, which resolves any ties. Judges confer in cases where there are extreme differences between their judgments; however, this recalibrating of scores is uncommon. The judging experience is an integral part of establishing and maintaining a judge’s level of certification. In other words, judges take the tasting very seriously, attending to those factors they are trained to discern, which results in very little discrepancy between judges on the score of a beer that they both rated, with the correlation between judges averaging 0.92. The top beers of each flight move
into a medal round where they are judged again in category flights. The highest ranking beers in medal round scoring are declared the winners of their respective categories.

This dataset is ideal for studying assimilation and contrast effects for hedonic experiences in a real-world setting because it contains multiple sets of sequential evaluation processes in random sequences (1,431 sequences over the eight years). Additionally, the judges are experts, trained to be observant for signs of satiation and fatigue (Strong and Piatz 2008). Finding assimilation and/or contrast effects in this data would speak to the pervasiveness and strength of these effects.

CONCEPTUAL DEVELOPMENT

In the literature on sequential evaluation, perhaps the most well-known result is Loewenstein and Prelec’s (1993) demonstration that people prefer improving sequences. These authors asked respondents to choose their preferred sequence among a series of events that included, for example, eating at home, eating at a fancy French restaurant, and eating an exquisite lobster dinner. Their conclusion is that people are farsighted and often wish to postpone better outcomes until the end. We should point out that all of their results are based on stated preferences for known sequences and not judgments of actual experiences. Raghunathan and Irwin (2001) found that for cases of domain match, improving (decreasing) sequences of the stimuli reduce (increase) the prospective happiness with the target, revealing that the nature of sequences might also induce contrast effects. The results are reversed for cases of domain mismatch. Like Loewenstein and Prelec (1993), their results were based on predicted preferences and not actual experiences. However, work by Novemsky and Ratner (2003) has cast doubt on

whether context effects, particularly hedonic contrast, documented in prospective and retrospective evaluations occur during actual experience. They find people recall contrast effects unless prompted to provide ratings of experiences during the consumption episode. Respondents’ evaluations of sequences of jelly beans reported while tasting them tended not to exhibit contrast. Also note that the above stream of research has looked at deliberately pre-arranged increasing or decreasing sequences; less is known about how hedonic events experienced in unanticipated, non-monotonic, or otherwise random sequences (such as in our data) are evaluated.

In terms of relative position in the sequence, work by Carpenter and Blackwood (1979) documented anomalous ratings for the first in a series. These authors had respondents rate various wildlife practices – for example, the acceptability of six different ways of killing coyotes (traps, slow poisons, etc.) on a 10-point scale. They find that when a method was presented first, it received either the highest or lowest rating and posited that the lack of an evaluative reference point before the first item was the reason for these extreme evaluations; prior items act as a “norm” for items appearing later. Similarly, Welch and Swift (1992), in tests of monadic sequences of four beverages, observe that the product in the first position receives a higher rating. Although they did not test for or document a mechanism, they posited that taster fatigue may be driving this effect. Other research suggests that stimuli characteristics help determine sequence effects. For example, Biswas, Grewal and Roggeveen (2010) find that for a pair of undesirable (desirable) experiential products, the option sampled first (second) is relatively preferred, and they attribute this to a recency effect leading to superior recall for the most recently sampled product. In a similar vein, de Bruin and Keren (2003) showed participants sequences of potential blind dates or dorm rooms such that each option either had unique positives or unique negatives. When choosing from a group of options with unique negatives, the
option presented first was most often preferred, but for a group of options with unique positives, the option presented last was most often preferred. Given that the goal of the judges in our data set is to compare each beer tasted to the “perfect beer” exemplar of that subcategory (i.e. they presumably start off with the “perfect beer” score of 50, and subsequently deduct points for flaws), they may be “primed” to look for differences and thus focus on negatives (the undesirable aspects). We therefore posit that the first beer may have a distinct advantage and be evaluated more favorably than successive beers. Stated more formally:

**H1:** Ratings of the stimulus in the first position will be higher on average than ratings of subsequent stimuli.

Based on the evidence suggesting that the natural tendency is to assimilate (Martin 1986; Martin, Seta, and Crelia 1990; Myers-Levy and Sternthal 1993) and that assimilation is a more spontaneous process, we expect assimilation effects to be prevalent in our data. In particular, we expect that the rating of the very first beer will act as an anchor (Tversky and Kahneman 1974; Chapman and Johnson 1999) and thus serve as a point of reference. In turn, we expect the remaining beers in the flight to assimilate to this value (Stapel and Koomen 1997; Stapel and Winkielman 1998). This leads to our second hypothesis:

**H2:** Across a sequence of trials, ratings assimilate to the first stimulus.

Note that hypothesis 2 implies assimilation effects would influence group or sequence means for a set of beers evaluated in a particular group or sequence. Observing assimilation effects
across sequences seems reasonable and probable because all of the beers within a flight belong to the same category, which brings similarities among the group to the forefront (Mussweiler 2003). However, this is not necessarily what happens between beers within a sequence. Stapel and Koomen (1997) predict that contrast effects should occur whenever context information provides a standard of comparison. In beer judging, each beer tasted is compared directly against every preceding beer in order to select a winner from each flight. Thus, it seems likely judges who are comparing individual beers relative to one another would contrast a beer they are tasting to other beers that came before it in the sequence.

In addition, if we consider the results of Myers-Levy and Tybout (1997) and Bickart (1993) as illustrative, we would expect to find contrast effects in our data. Our respondents (expert judges in a national competition) are arguably in a deliberative mode of evaluation, as opposed to casual or spontaneous, and both possess and are willing to expend substantial cognitive resources in making their evaluation. Further, for a beer that is especially good or especially bad (extremes) we expect to see contrast effects (Herr 1986; Herr, Sherman, and Fazio 1983). Accordingly, we expect to see evaluations of a focal beer contrasted to extremely good preceding beers that judges may be holding as the standard for future beers to beat, and to extremely bad beers that have appeared earlier in the flight. This leads to our third hypothesis:

**H3a:** Ratings within a sequence will contrast to adjacent stimuli.

**H3b:** Ratings within a sequence will contrast to extreme stimuli.
It is important to point out that documenting the effects predicted by Hypotheses 2 and 3 would be the first demonstration of the coexistence of assimilation and contrast effects within the same sequence of hedonic experiences.

Subcategory effects and domain match is another area in the literature that sees conflicting results. The general conclusion regarding context effects and domain match/mismatch in psychophysics and tasks of impression-formation is that assimilation to the context occurs when the target and context are viewed to be similar, while contrast occurs when the target and context are viewed to be dissimilar (Damisch, Mussweiler, and Plessner 2006; Myers-Levy and Sternthal 1993; Schwarz and Bless 1992). However, for hedonic contrast, the results appear to be reversed such that contrast occurs when the stimulus and target belong to the same category, and assimilation occurs when they belong to different categories (Fechner 1898; Raghunathan and Irwin 2001; Zellner et al. 2002). Because our data consist of sequences of hedonic judgments, the results of this second stream of literature hold more relevance. In our data, all beers within a flight belong to the same category (e.g. Porter), yet within that category, there are frequently beers from two or more “sub-categories” (e.g. Brown Porter, Robust Porter, and Baltic Porter). For the first beer of a new subcategory, the preceding beer, which was the last beer of the previous subcategory, might be considered to be a stimulus from a “different domain.” Hence, we hypothesize that contrast effects, if present, will be stronger within a subcategory and may be diminished across subcategories.

**H4:** Ratings will contrast more strongly with stimuli within the same subcategory.

In order to test our hypotheses, we needed to better understand the data and rule out potential confounds. We would then be able to develop a model that controls for different factors
such as satiation and beer style while isolating contrast effects from the effects of assimilation. Unlike much of the previous work in this area that has documented these effects utilizing separate tests, we set out to test for these effects simultaneously. In what follows, we present our initial analysis of the data in which we assess possible trends and other data characteristics prior to the development of our formal model.

**PRELIMINARY DATA ANALYSIS**

Each record in our data contains a unique beer ID, the year, name and ID of the brewer, the specific category and subcategory of the beer, the round (Round 1 or Round 2) in which it was judged, the flight number, date, the name and certification level of the two (or three) judges who rated the beer, the placement of that beer in that particular round (only if it was in the top three), the position of the beer in the flight (order), and the total points out of 50 awarded to the beer by each of the judges who rated it. Prior to formally investigating our hypotheses, we conducted a preliminary analysis of the data in order to test for anomalies (e.g. differences across rounds, years, or beer types) and any systematic judging patterns that may need to be accounted for in the main model. For example, judges may tend to assign scores in a much narrower range (central scoring) as time progresses simply because palate fatigue causes the beers to taste more and more similar over time. In addition, as judges become fatigued during long tasting flights, they may assign the darker beers that occur later in the flight much higher scores than earlier beers simply because they stand out as being much more flavorful (positive drift). Alternately, preference for lighter beers or even satiation could affect the hedonic enjoyment of later beers and result in judges assigning progressively lower scores to beers as time progresses (negative
In order to more accurately measure the effects of interest (contrast and assimilation), the model should account for these types of scoring patterns if they are present (Wolfe 1996; Wolfe and Wolfe 1997).

Differences across Rounds, Years, and Types of Beers

Overall, scores in both rounds are widely dispersed, with values ranging from 10.5 to 49.5 in the preliminary round (Round 1) and from 16.0 to 48.5 in the medal round (Round 2). The average score in Round 1 is 32.4. The Round 2 beers are comprised of the top beers in Round 1 and, not surprisingly, the average Round 2 score is higher at 33.2 ($p < 0.01$). However, this Round 2 average is lower than the average score of the very same beers when they were judged in Round 1 ($M = 38.6, p < .01$). This difference in the medal round scores can be attributed to the judges’ natural scale recalibration in order to prevent ceiling effects. In addition, we expected there to be less variation for scores in Round 2 because the tail consisting of all the really “bad” beers has been eliminated ($\sigma^2_{\text{Round1}} = 39.7$ vs. $\sigma^2_{\text{Round2}} = 30.2$; $F = 1.28, p < .01$). It is critical to point out that a sizable percentage of winning beers (33%) are decided by one (1) point or less and that a difference of up to two (2) points decides 58% of the winners in our data. Hence, small changes in evaluations due to assimilation or contrast effects can have a very meaningful impact on outcomes, and in this case, one of the largest homebrew competitions in the U.S.

______________________________

Insert table 2 about here

______________________________
There appears to be a slight upward trend for average scores of beers across years. On average, scores increase by 0.12 points every year ($p < .01$; see Table 2). This pattern was discussed with the organizers of the competition, and this yearly drift is believed to be caused by the greater craftsmanship and expertise exhibited by the participating brewers as time progressed.

Mean scores for 107 sub-categories of beer range from 29.4 to 37.3 ($F = 2.56; p < .01$), showing that certain subcategories on average receive much higher scores than other subcategories. A follow-up interview conducted with the competition organizers and beer judges revealed why certain subcategories may receive higher scores on average. Organizers stated that the list of higher scoring beers agreed with their a priori beliefs that certain beer styles are more difficult to produce (e.g. Classic American Pilsner and Gueuze/Geuze-Style Ale) and only the more proficient brewers tend to enter beers in those categories. The subcategories with higher mean scores consist of both light and dark beers and therefore appear both early and late in the flight sequences. In other words, while some subcategories stand out, there is no systematic pattern regarding where these subcategories appear within flights.

Central Scoring

If palate fatigue, or a decline in the ability of judges to discriminate as they consume more and more beer, existed in our dataset, we would expect the variance in scores to decrease with flight position (Wolfe and Wolfe 1997). However, a regression of flight position on score standard errors, a test that ignores the uncertainty in the standard error estimates and thus is more
likely to yield test results deemed to be “statistically significant,” reveals an increasing trend (+.02; \( p < .01 \)). Consequently, we observe no evidence supporting palate fatigue.

Positive Drift and Negative Drift

Although beers are randomly assigned to flights, the order of beers within the flight is not completely random. Judges are advised to taste beers that belong to lighter, less-flavorful subcategories before beers belonging to stronger, more-flavorful sub-styles. This is done to diminish the potential effects of palate fatigue as well as help prevent stronger beers from overpowering the flavor distinctions of milder beers. The systematic sub-flight ordering could influence the scoring in two potential ways – if heavier beers are preferred over lighter beers, then beers occurring in sub-categories appearing later in the flight would receive higher ratings. This preference for more extreme flavors could manifest in a positive drift of overall scores as the flight progresses. On the other hand, if lighter beers are preferred over stronger beers, then beers occurring earlier in the flight would receive higher ratings, manifested in a negative drift of overall scores as the flight progresses.

In addition, there is another potential source for a systematic negative drift in scores appearing in the data. If judges satiate as they taste each beer (Coombs and Avrunin 1977; Rolls et al. 1981, 1984), each subsequent beer should be less enjoyable than the prior, and we would expect a decrease in a flight’s average ratings as the length of the flight increases. Further, we felt it was necessary to test for satiation at both the flight and sub-flight (beers belonging to the same subcategory within a flight) level as changes in subcategories of beer may be significant enough for judges to combat satiation (Redden 2008).
The average ratings by position within a flight and within a subcategory are shown in Table 3. The means do not reveal any discernible positive or negative drift or trends in our data at the flight or sub-flight level. Additionally, a regression of flight position on scores is not significant. We therefore find no evidence of satiation or systematic negative or positive drift in scores down the flight.

---

Insert table 3 about here

---

**EMPIRICAL ANALYSIS**

Model

Actual beer quality is expected to vary across entries and across years in a fashion similar to what has been found with respect to wine quality (Ashenfelter, Ashmore, and Lalonde 1995). Because beers are randomly ordered within their respective subcategory structure, objective characteristics that affect beer quality (e.g. ingredient quality or brewing technique) would not be associated with sequence effects. Our model, which focuses on beer scores as a function of their order within a sequence, therefore measures what is of primary interest: the scoring behavior of judges. All analyses were done at the flight level. In order to simultaneously test the effects of preceding beers and extremes such as the running maximum and minimum scores, our model effectively estimates coefficients for beers in the third position and later in the sequence. Hence,
we are unable to test Hypothesis 1 within the purview of the model, and therefore we proceed by first testing it separately.

*Sequence Position Effects.* In order to test Hypothesis 1, we compare the average score of the beers in the first position to the average score of all beers in subsequent positions. As expected, the mean of the ratings of beers in the first position of the sequence, $M_{\text{position1}}$, is rated significantly higher than the mean of the ratings of beers in subsequent positions, $M_{\text{rest}}$ ($M_{\text{position1}} = 33.5; M_{\text{rest}} = 32.4; p < .01$), supporting Hypothesis 1.

*Hierarchical Bayesian Random Effects Model.* Our data comprise more than 5,000 beers belonging to over 100 style subcategories, assigned to and evaluated in 1,431 judge-flight sequences. Our hypotheses involve variables at two levels: an individual entry or beer in a particular position in the flight, and the flight (judge-flight combination) to which it is assigned, which is the level at which the model is estimated. Given the multilevel nature of our data, and in order to allow for the most flexibility in the error structure across levels, we formulate and estimate a flexible hierarchical linear Bayesian random parameters regression model of the following nature: $y_i = X_i \beta_i + \epsilon_i; \epsilon_i \sim \text{iid } N(0, \sigma_i^2)$. Readers interested in further details of the estimation procedure are referred to Allenby and Rossi (1999) and Gelfand and Smith (1990).

This model represents a set of independent regression equations with a different error variance for each equation. The equations are tied together through a common prior distribution on the $\{\beta_i\}$ as follows. Specifically, in our model, apart from flight-specific heterogeneity, within a flight we account for beer style or subcategory-specific, as well as judge-specific heterogeneity. We consider beer $k; k \in (1...5,244)$ in flight $i; i \in (1...688)$; evaluated by judge $j; j$
ε (1…3) across the eight years. We therefore use $y_{ijk}$ to represent the score (out of 50) of the $k^{th}$ beer in the $i^{th}$ flight as assigned by the $j^{th}$ judge. This is the dependent measure for the model.

The conditional distribution of $y_{ijk}$ given a vector of predictor variables may be written as:

$$y_{ijk} = \beta_0 + \beta_1 y_{ij,k-1} + \beta_2 \text{FIRST}_{ij} + \beta_3 \text{MAX}_{ijk} + \beta_4 \text{MIN}_{ijk} + \beta_5 \text{SUBCATCHANGE}_{ik} + \beta_6 \text{SUBCATCHANGE}_{ik} y_{ij,k-1} + \beta_7 \text{STYLE}_{ik} + \epsilon_{ijk}$$

where

$$\beta_i \sim N(\beta, V_\beta)$$

represents a normal prior with a common mean vector and covariance matrix $V_\beta$. Note that the vector of coefficients $\beta_i$ represents flight-level random effects while the intercept term $\beta_0$ would be flight-level intercepts, representing flight-level assignment effects. Equation (2) denotes the hierarchical structure of the model.

The variables in the model are detailed below, where Table 4 provides a succinct summary of the variables included in the model.

---

Insert table 4 about here

---

First, $y_{ij,k-1}$ represents the score of the preceding beer in the flight; i.e. the $k-1^{th}$ beer in the $i^{th}$ flight as assigned by the $j^{th}$ judge. This would test for contrast among adjacent stimuli (hypothesis 3a). Note that a positive coefficient would reflect positive correlation, indicating assimilation, while a negative coefficient would reflect negative correlation, indicating contrast. For example, if the sign on $\beta_{1i}$, the parameter for score of the previous beer, is negative, it would mean that the score of the focal beer contrasts to this value. A higher (lower) the score of the previous beer would indicate a lower (higher) score for the focal beer. We include the score of
the first beer in the \(i^{th}\) flight as assigned by the \(j^{th}\) judge directly as an independent variable (FIRST\(_{ij}\)) in order to test Hypothesis 2 and assess whether the first beer impacts scores for the rest of the beers in the flight (i.e. an anchoring or assimilation effect). This would be true if the sign on \(\beta_{2i}\) were positive; indicating that the higher (lower) the score of the first beer, the higher (lower) would be the score of the focal beer. We also include the maximum score leading up to the \(k^{th}\) beer in the \(i^{th}\) flight as assigned by the \(j^{th}\) judge as MAX\(_{ijk}\); this is to test whether an “extremely high” value would have an impact on subsequent evaluations. We analogously define MIN\(_{ijk}\) for the minimum score leading up to the focal beer. For an example of contrast to extremes, as predicted by Hypothesis 3b, if \(\beta_{4i}\), the coefficient on MIN\(_{ijk}\), were negative, the lower the score of the minimum beer, the higher would be the expected score of the focal beer. Taken together, these variables enable us to test for assimilation and contrast effects between extreme and adjacent beers.

We define a dummy variable (SUBCATCHANGE\(_{ik}\)) that equals 1 if the preceding beer (the beer just before the \(k^{th}\) beer in the \(i^{th}\) flight) was in a different subcategory than the focal beer. In order to determine whether context effects are enhanced or diminished at the points of subcategory change and thus test Hypothesis 4, we include the interaction term of the SUBCATCHANGE\(_{ik}\) dummy with the score of the previous beer. The interpretation of this interaction term will be subject to the sign of the parameter estimates for the main effect. For example, if \(\beta_{1i}\), the parameter for score of the previous beer is negative (contrast effect), and if the coefficient of the interaction term \((\beta_{6i})\) is positive, this would imply that a subcategory change (SUBCATCHANGE\(_{ik}\) = 1) decreases the intensity of the contrast effect. In other words, contrast effects may be exacerbated by domain match and alleviated by domain mismatch, as hypothesized. Finally, to allow for the subcategory differences evident from our preliminary data
analysis, we include 107 subcategory dummy variables (STYLE\textsubscript{ik}) that indicate the style subcategory to which the k\textsuperscript{th} beer in the i\textsuperscript{th} flight belongs.

Because our analysis is within flights rather than across time, parameter estimates will not be influenced by annual trends. Consequently, we do not include any variables in the model to control for the annual improvement trend pointed out earlier.

In the data, 5,244 individual beers (only considering positions 3 onwards in flights) were evaluated by 2-3 judges each, resulting in 10,928 individual beer evaluations by individual judges, and 1,431 judge-flight sequences. The regression in equation (1) is therefore estimated 1,431 times for 1,431 groups of beers, each representing a particular flight-judge combination.

Following Rossi, Allenby and McCulloch (2005), the priors on the collection of $\beta_i$ (equation 2) are specified through a two-stage process. First, a normal prior (with a common mean vector $\bar{\beta}$ and a fixed variance matrix $V_\beta$) is specified on $\beta_i$ (equation 2) and then a second-stage prior is specified on the parameters of this distribution (equations 3-4). The prior and posterior for the covariance matrix of the multivariate normal distribution are of the inverted Wishart form (equation 4); $V_\beta$ represents the unobserved heterogeneity, with $V$ characterizing the extent of the heterogeneity. We also assume the standard natural conjugate prior (inverse gamma) on the regression error variance, which specifies each of the error variances to be independent (equation 5). Note that the second stage priors are set to be proper but very diffuse ($A = 0.01I; \nu_i = 3; \nu = k$ (number of variables) + 3; $V = \nu^*I_k$). Since we have a large number of units in the analysis, the data are expected to overwhelm these priors. Note that inferences about specific flights may be obtained from the posterior distribution of $\beta_i$.

(2) $\beta_i \sim \text{N}(\bar{\beta}, V_\beta)$

(3) $\bar{\beta} \sim \text{N}(0, A^{-1})$
(4) $V_{\beta} \sim IW(\nu, V)$

(5) $\sigma_i^2 \sim \frac{\nu_i s_i^2}{X_{V_i}^2}$

The algorithm for estimating the model parameters is a Monte Carlo Markov Chain (a Gibbs sampler) that draws recursively from the posterior conditional distribution of the model parameters. This is a well-accepted estimation technique for estimating complex hierarchical models (Gelfand and Smith 1990; Allenby and Ginter 1995). To complete one iteration of the chain, draws from the following conditional distributions are obtained and used as conditioning arguments in subsequent draws:

(6) $y_{ijk} | X_{ijk}, \beta_i, \sigma_i^2$

(7) $\beta_i | \bar{\beta}, V_{\beta}$

(8) $\sigma_i^2 | v_i, s_{0i}^2$

(9) $V_{\beta} | \nu, V$

(10) $\bar{\beta} | V_{\beta}, A$

The Gibbs sampler proceeds by cycling though equations 6 through 10 until draws from the distributions converge to a stationary distribution. Statistics of interest are obtained empirically by calculating the appropriate sample statistics from the draws. In the subsequent analysis, a total of 50,000 draws were used to estimate the model with initial parameter values set to zero. Time series plots of the draws indicate that the distributions converged to the stationary distributions after about 10,000 iterations, which was used as burn-in for the rest of the analyses. Convergence was assessed using the Heidelberger and Welch Convergence Diagnostic as implemented by the BOA program in R.
Results

Contrast Effects. Analysis of the marginal posteriors for score of the previous beer (posterior mean -0.07; Figure 1a; Table 5a), the score of the running maximum (posterior mean -0.55) and score of the running minimum (posterior mean -0.50) reveal that >99% of the mass is below zero, revealing that the dependent variable is negatively correlated with these measures. The score of the focal beer being negatively correlated with the score of the predecessor beer indicates contrast effects between adjacent beers. In other words, the higher (lower) the score of the previous beer, the lower (higher) the score we would expect for the focal beer, supporting hypothesis 3a. The negative correlation with the running maximum and minimum scores indicates contrast effects against extreme scores. The higher (lower) the score of the running maximum/minimum, the lower (higher) the score we would expect for the focal beer, supporting hypothesis 3b. Thus we see evidence of contrast effects between adjacent stimuli as well as against extreme stimuli.

Assimilation effects. The marginal posterior for the score of the first beer (posterior mean 0.65; figure 1b) is positive; the score of the focal beer is positively correlated with the score of the first beer of the flight, indicating assimilation or anchoring effects. In other words, the higher (lower) the score of the first beer, the higher (lower) we would expect the score to be of the focal
beer, supporting hypothesis 2. Moreover, the multi-modal posterior distribution of the flight-level intercept estimates (figure 1b) is consistent with the existence of assimilation effects at the level of a flight (95% confidence interval; flight means 8.52 - 52.3). These results indicate that assimilation or anchoring effects influence all beers in a particular sequence being evaluated together, effectively influencing group or sequence means.

---

Insert figure 1b about here

---

*Subcategory effects.* The interaction term of the score of the previous beer and the subcategory change dummy variable is positive (posterior mean .17). Given that the distribution on the score of the previous is negative (contrast effect), the positive sign on the interaction term indicates an attenuation of the contrast effect between adjacent beers at the point of a subcategory change. In other words, scores contrast more strongly within beers of the same subcategory style, thus supporting hypothesis 4. Also, the marginal posterior for the subcategory change dummy variable is negative (posterior mean -4.74; figure 1c), indicating that the score of the first beer in a new sub-flight is adjusted downward. While we have no a priori theory for this effect, it could be explained by a possible recalibration of the scale by the judges at the first position of a new subcategory style of beer within the flight.

---

Insert figure 1c about here

---
**Beer style category effects.** As expected, there are significant differences in scores as per the main style or category of beer. Table 5b shows the posterior estimates of the significant style categories of beer, indicating that certain styles like European Amber Lager receive higher scores, while others like Light Scottish Ales receive lower scores on average.

---

**Discussion.** The hierarchical Bayesian framework allows us to investigate multiple effects at different levels of the data (across and within flight) simultaneously in one estimation procedure. The estimation reveals assimilation or grouping effects at the level of a flight of beers being evaluated together, and specifically assimilation/anchoring effects to the score of the first beer of the flight. We also find contrast effects within the flight—between adjacent beers, as well as to extremes (running maximum and minimum scores). In summary, we are able to test for and thus document separate assimilation and contrast effects in the same sequence of hedonic experiences.
GENERAL DISCUSSION

In this research, we successfully isolate both assimilation and contrast effects in the same sequence of hedonic experiences. The fact that we can identify and separate simultaneous assimilation effects and contrast effects in our dataset is unique with respect to the extant literature. For example, while beer scores within a flight assimilated to each other, beers evaluated in the flight contrasted against adjacent and extreme scores as well, and these effects are measured simultaneously by the hierarchical model. We do not believe that this is an artifact of our data but posit that previous researchers either did not look for simultaneous effects or did not have data that would allow them to uncover such effects. Our analysis reveals that at the overall flight level, assimilation effects are evident and beers in the same flight (particularly the first beer) appear to serve as frames of reference for each other. However, within the flight, contrast effects are also prominent between adjacent beers as well as after extremes within the sequence. To the best of our knowledge, this is the first time that these two effects have been shown to coexist and been isolated in the same dataset.

Our results are important for marketing managers and consumers alike. Recall we pointed out that at the Bluebonnet Brew-off, one of the largest homebrew competitions in the U.S., for 33% of the flight contests, the first place beers won by one (1) point or less, and a difference of up to two points decides 58% of the winners. Given the substantial effects we find in our model, these statistics reveal that a sizeable proportion of the outcomes may be influenced by assimilation and contrast effects. Further, it is well known that experiences judged jointly provide different outcomes than evaluations made separately (Hsee et al. 1999). The conventional wisdom is that sequential evaluation of randomly ordered items provides more
objective measures of comparative quality and thus is preferred when assessing preferences and judging products. Our results contribute to a growing literature documenting the psychological ramifications inherent in how one constructs choice sets or sequences for trial. Unbeknownst to many wine enthusiasts or even those of us who believe we randomly begin watching one specific TV show before flipping channels, what we have experienced and the order in which we experience it is impacting our evaluation in numerous subtle ways. Our research reconfirms that judging one experience can unduly influence our judgment of subsequent events and thus “color” the entire sequence of experiences. Our results also highlight the fact that randomization is not the panacea it appears to be.

Consider an experience of tasting successive samples randomly in a food court in a mall. While one particularly tasty sample may improve the average evaluation of the food as a whole (assimilation), it may also create strong contrast effects for the samples tasted directly following that extreme in quality. In other words, while the entire experience at that mall may compare favorably against a sampling experience at another mall (or at the same mall on another day), individual samples tasted that day may be rated above or below their objective quality due to within-sequence contrast effects.

Or consider a consumer shopping experience with more profound implications. One author went to test drive the new 2009 Jaguar XF. He began with the “Supercharged” top-of-the-line version before progressing downward by subsequently driving the “Luxury” version and concluded by moving up and driving the “Premium Luxury” version. His perception of Jaguar’s new sport sedan was likely influenced by the first car he drove, the Supercharged model that comes equipped with all of the premium features standard. Consequently, the Jaguar XF in general (assimilation) was very well-regarded. It compared extremely favorably to the BMW 5
series and Mercedes E-class. However, while evaluating the experience of driving each model FX, the difference between the 420 horsepower Supercharged model and the 300 horsepower luxury versions became evident, as did the contrast between the ultra-premium Bowers and Wilkins 440-watt stereo and the Jaguar 320-watt Jaguar-branded standard stereo. In other words, within the sequence, contrast effects emerged. Like the results of our model, the evaluations that emerge from test driving different models within a new line of vehicles is likely to be impacted by both assimilation and contrast simultaneously. In addition, to fully comprehend the effect of different features, it would be necessary to first control for assimilation.

This begs the question of what marketing managers might do to diminish these effects. One suggestion may be to encourage repeat trials in new sequence orders. In the Bluebonnet Brew-off, each judge could sample a flight multiple times in different sequences, or, more cost-effectively, the two judges could each sample the same flight in a different order. Similarly, after driving the luxury model, the author could try the supercharged version again to see if it really feels that much faster. While we have no evidence to confirm that revisiting the experiences in a different order would moderate the assimilation and contrast effects we observed (it may create a number of new effects), we suspect that a greater sampling both in number and duration would help judges form more objective evaluations.

In this research, we show that individuals’ evaluations of sequences of hedonic experiences exhibit contrast effects (thus empirically demonstrating real-time, experienced hedonic contrast) as well as assimilation effects simultaneously. While our research reveals how individual experiences impact each other, as mentioned above, we neither posit nor test methods to attenuate or eliminate these effects. One expects, however, normatively speaking, that marketers and consumers would prefer to know how an experience rates independent of context.
Future research should investigate methods that either measure or control these effects such that practitioners, be they the firm or judges or even consumers might remove them to obtain “purer” measures.

In addition, while many of the contrast and assimilation effects documented in the literature have been ascertained without simultaneous tests, it may be worth selectively re-investigating certain phenomenon to see whether the effects withstand such testing. We do not call any specific effects into question, as we do not have any reason to expect specific results to change. However, it may be the case that researchers who did not observe contrast effects may have if they had accounted for assimilation, and those who did observe contrast effects may have also observed assimilation, which might have suppressed the size of the contrast effect. Our hope is that in the future, researchers who test for these types of effects utilize methodologies that allow them to test for both simultaneously. One obvious limitation of our work is that our analysis is constrained to a single data set. While the data are rich and appropriate for such testing, we cannot speak to whether researchers might find similar patterns of effects in other sequences of hedonic experiences. We believe other researchers reflecting upon this work will be inspired to test for the simultaneity of these effects.
Appendix: Beer Scoring Sheet Used by Judges; Note Scoring Guide on Bottom Left

# Bluebonnet Brew-off 2006
19th Annual—March 17-18, 2006

## Beer Score Sheet

<table>
<thead>
<tr>
<th>Judge Name (print):</th>
<th>Judge BJCP ID:</th>
</tr>
</thead>
</table>

### Judge Qualifications/BJCP Rank:
- Apprentice
- Recognized
- Certified
- National
- Master
- Honorary Master
- Professional Brewer
- Novice (non-BJCP)
- Experienced (but not in BJCP)

### Descriptor Definitions (Mark all that apply):
- Acetaldehyde - Green apple-like aroma and flavor.
- Alcoholic - The aroma, flavor, and warming effect of ethanol and higher alcohols. Sometimes described as "hot."
- Astringent - Puckering, lingering harshness and/or Dryness in the finish/bitter taste; harsh graininess; huskiness.
- Butty - Artificial butter, butterscotch, or toffee.
- Aroma and/or flavor of any ester (fruits, fruit flavors, or roses.)
- Crushy - Aroma/flavor of fresh-cut grass or green leaves.
- Light-Struck - Similar to the aroma of a skim. Aroma and/or flavor of any ester (fruits, fruit flavors, or roses.)
- Metallic - Tinny, colly, copper, iron, or blood-like flavor.
- Musty - Stale, musty, or moldy aromas/flavors.
- Oxidized - Any one or combination of wod/vinous, cardboard, papery, or sherry-like aromas and flavors.
- Phenolic - Spicy (clove, pepper), smoky, plastic, plastic adhesive strip, and/or medicinal (chlorophenolic).
- Solvent - Aromas and flavors of higher alcohols (fused alcohols). Similar to acetone or lacquer thinner aromas.
- Sour/Acidic - Tartness in aroma and flavor. Can be sharp and clean (lactic acid), or vinegar-like (acetic acid).
- Sulfur - The aroma of rotten eggs or burning matches.
- Vegetal - Cooked, canned, or rotten vegetable aroma.
- And flavor (cabbage, onion, celery, asparagus, etc.)
- Yeasty - A bready, sulfury or yeast-like aroma or flavor.

### Special Ingredients:
- Bottle Inspection: ☐ Appropriate size, cap, fill level, label removal, etc.

### Comments:
- Aroma (as appropriate for style)
- Flavor (as appropriate for style)
- Appearance (as appropriate for style)
- Mouthfeel (as appropriate for style)
- Overall Impression

### Scoring Guide

<table>
<thead>
<tr>
<th>Overall Impression</th>
<th>0-13</th>
<th>14-20</th>
<th>21-25</th>
<th>26-30</th>
<th>31-35</th>
<th>36-40</th>
<th>41-45</th>
<th>46-50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intolerable</td>
<td>(0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>(1-2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fairly poor</td>
<td>(3-5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorly acceptable</td>
<td>(6-10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acceptable</td>
<td>(11-15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above average</td>
<td>(16-20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>(21-25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very good</td>
<td>(26-30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td>(31-35)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outstanding</td>
<td>(36-40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Stylitic Accuracy

- Classic Example: ☐ ☐ ☐ ☐ ☐ ☐ Not to Style
- Fitness: ☐ ☐ ☐ ☐ ☐ ☐ Significant Flaws
- Intellligence: ☐ ☐ ☐ ☐ ☐ ☐ Lifeless

### Subcategory (spell out)

### Entry # | Category # | Subcategory (a-d)
REFERENCES


Journal of Econometrics, 89 (1–2), 57–78.


Table 1: Summary Statistics of Flight and Sub-flight Length

<table>
<thead>
<tr>
<th></th>
<th>Total number</th>
<th>Mean length</th>
<th>Modal length</th>
<th>Minimum length</th>
<th>Maximum length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flights</td>
<td>688</td>
<td>9.6</td>
<td>10</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Sub-flights</td>
<td>1756</td>
<td>3.8</td>
<td>2</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>
Table 2: Average Ratings (Standard Deviations) by Year and Round

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round 1</td>
<td>32.0 (6.7)</td>
<td>32.5 (7.0)</td>
<td>31.9 (6.0)</td>
<td>31.9 (6.0)</td>
<td>32.3 (6.3)</td>
<td>32.3 (5.6)</td>
<td>32.9 (5.8)</td>
<td>33.1 (6.2)</td>
<td>32.4 (6.3)</td>
</tr>
<tr>
<td>Round 2</td>
<td>31.5 (6.3)</td>
<td>33.6 (5.5)</td>
<td>33.6 (5.2)</td>
<td>33.6 (5.3)</td>
<td>34.4 (5.6)</td>
<td>32.9 (5.3)</td>
<td>33.0 (5.1)</td>
<td>33.1 (5.4)</td>
<td>33.2 (5.5)</td>
</tr>
<tr>
<td>Average</td>
<td>31.9 (6.6)</td>
<td>32.7 (6.7)</td>
<td>32.3 (5.9)</td>
<td>32.3 (5.9)</td>
<td>32.8 (6.2)</td>
<td>32.5 (5.6)</td>
<td>32.9 (5.6)</td>
<td>33.1 (6.0)</td>
<td>32.5 (6.1)</td>
</tr>
</tbody>
</table>
Table 3: Average Rating (Standard Deviation) by Position in Flight and Subflight

<table>
<thead>
<tr>
<th>Position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flight</td>
<td>33.5</td>
<td>32.1</td>
<td>32.0</td>
<td>32.3</td>
<td>32.1</td>
<td>32.6</td>
<td>32.8</td>
<td>33.0</td>
<td>32.9</td>
<td>32.7</td>
<td>32.1</td>
<td>32.3</td>
</tr>
<tr>
<td>Sub-flight</td>
<td>33.4</td>
<td>32.6</td>
<td>32.2</td>
<td>32.1</td>
<td>31.7</td>
<td>32.3</td>
<td>32.2</td>
<td>32.4</td>
<td>32.6</td>
<td>32.5</td>
<td>30.4</td>
<td>29.1</td>
</tr>
</tbody>
</table>
### Table 4: Variables in model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Interpretation/reason for inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_{ijk}$</td>
<td>Continuous; score of an entry; range 10.5-49.5</td>
<td>Dependent variable</td>
</tr>
<tr>
<td>$y_{ij,k-1}$</td>
<td>Continuous; score of predecessor beer</td>
<td>Test Hypothesis 3a; whether ratings contrast between adjacent beers</td>
</tr>
<tr>
<td>FIRST</td>
<td>Continuous; score of first beer of flight</td>
<td>Test Hypothesis 2; whether ratings assimilate to first beer score</td>
</tr>
<tr>
<td>MAX</td>
<td>Continuous; score of running maximum leading up to focal beer</td>
<td>Test Hypothesis 3b; whether there is contrast effect versus extremes</td>
</tr>
<tr>
<td>MIN</td>
<td>Continuous; score of running minimum leading up to focal beer</td>
<td>Test Hypothesis 3b; whether there is contrast effect versus extremes</td>
</tr>
<tr>
<td>STYLE</td>
<td>Dummy variables; to account for 107 (subcategory) styles</td>
<td>Control for subcategory differences evident from preliminary data analysis</td>
</tr>
<tr>
<td>SUBCATCHANGE</td>
<td>Dummy variable = 1 if preceding beer in different subcategory; 0 otherwise</td>
<td>Control for subcategory change effects</td>
</tr>
<tr>
<td>$y_{ij,k-1} \times$ SUBCATCHANGE</td>
<td>Interaction of score of predecessor beer and subcategory change dummy</td>
<td>Test Hypothesis 4; whether contrast effects are diminished or enhanced by subcategory changes</td>
</tr>
</tbody>
</table>
Table 5a: Model Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variable</th>
<th>95% Posterior Interval</th>
<th>Posterior estimate (standard deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{0i}$</td>
<td>Intercept</td>
<td>(8.52, 52.3)</td>
<td>46.57 (2.48) **</td>
</tr>
<tr>
<td>$\beta_{1i}$</td>
<td>Score of previous beer</td>
<td>(-.11, .03)</td>
<td>-.07 (.02) **</td>
</tr>
<tr>
<td>$\beta_{2i}$</td>
<td>Score of first beer</td>
<td>(.49, 1.62)</td>
<td>0.65 (.08) **</td>
</tr>
<tr>
<td>$\beta_{3i}$</td>
<td>Score of running maximum</td>
<td>(-.63, -.44)</td>
<td>-.55 (.04) **</td>
</tr>
<tr>
<td>$\beta_{4i}$</td>
<td>Score of running minimum</td>
<td>(-.58, -.39)</td>
<td>-.50 (.04) **</td>
</tr>
<tr>
<td>$\beta_{5i}$</td>
<td>Subcategory change dummy</td>
<td>(-6.92, -1.45)</td>
<td>-4.74 (1.18) **</td>
</tr>
<tr>
<td>$\beta_{6i}$</td>
<td>Interaction sc dummy and score of previous beer</td>
<td>(.07, .24)</td>
<td>.17 (.04) **</td>
</tr>
</tbody>
</table>

\* interpreted as: the posterior probability that the parameter $\beta_{0i}$ lies in the interval 8.52 to 52.3 is 0.95.

** significant at $p < .01$;
Table 5b: Model Parameter Estimates; Beer style effects

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>sc48 European Amber Lager</td>
<td>Octoberfest</td>
<td>6.42**</td>
<td>2.15</td>
</tr>
<tr>
<td>sc33 Brown Ale</td>
<td>Mild</td>
<td>-4.69**</td>
<td>1.66</td>
</tr>
<tr>
<td>sc49 European Amber Lager</td>
<td>Vienna</td>
<td>5.51**</td>
<td>2.09</td>
</tr>
<tr>
<td>sc104 Wheat Beer</td>
<td>Bavarian Dunkelweizen</td>
<td>-3.75**</td>
<td>1.84</td>
</tr>
<tr>
<td>sc79 Light Lager</td>
<td>Standard American Lager</td>
<td>6.18*</td>
<td>3.21</td>
</tr>
<tr>
<td>sc62 Koelsch and Altbier</td>
<td>Northern German Altbier</td>
<td>4.83*</td>
<td>2.53</td>
</tr>
<tr>
<td>sc82 Pilsner</td>
<td>German Pilsner</td>
<td>4.36*</td>
<td>2.30</td>
</tr>
<tr>
<td>sc60 Koelsch and Altbier</td>
<td>Duesseldorf Altbier</td>
<td>4.65*</td>
<td>2.47</td>
</tr>
<tr>
<td>sc88 Scottish Ales</td>
<td>Light 60</td>
<td>-8.51*</td>
<td>4.81</td>
</tr>
<tr>
<td>sc24 Bitter and English Pale Ale</td>
<td>Ordinary Bitter</td>
<td>-3.77*</td>
<td>2.19</td>
</tr>
</tbody>
</table>

**p < .05; *p < .1
Figure 1a: Marginal posteriors (PREVIOUS, MAX and MIN); Contrast effects
Figure 1b: Marginal posteriors (FIRST and INTERCEPT); Assimilation effects
Figure 1c: Marginal posteriors (SUBCAT CHANGE and SC*PREVIOUS); Subcategory effects