Is Beer Beauty in the Eye of the Beholder? Ratings of Professionals and Nonprofessionals

Grace Allred, Nina Filippini, and Duane Lundy
Faculty Mentor: Duane Lundy, Ph. D.

Abstract

This archival study investigated appraisers of beer aesthetic quality. In total, over 700 beers were randomly selected from beer critic books that contained quantitative ratings, and critics' ratings were analyzed in terms of histograms and related statistics. We compared beers from professional beer critics' books, professional websites and nonprofessional consumer websites (N = 2,200 ratings). It was predicted that professionals would have positive correlations in their ratings, with ratings showing mound-shaped distributions of ratings approaching normality. Nonprofessionals, in contrast, were predicted to be less positively correlated with each other and to show less normalized distributions. Support for the hypotheses was mixed.

Keywords

aesthetic consensus, aesthetic merit, aesthetic quality, rating distributions, statistical analysis of beer judgments

---

This study is about the aesthetics of beer. Aesthetics is the branch of philosophy concerned specifically with taste. Aesthetics has been defined as "[t]he study of the feelings, concepts, and judgments arising from our appreciation of the arts or of the wider class of objects considered moving, beautiful, or sublime" (Blackburn, 1996, p. 8). Different areas of aesthetics can be studied using parallel methodologies. Some aesthetics studies in the past have shown: 1) positive correlations in ratings among professional music and film critics, suggesting they agree more than disagree about quality (Boor, 1990; Lundy, 2010; Simonton, 2011), and 2) a mound-shaped distribution of ratings, approaching normality, implying that most works are near neutral in quality and extremes of quality are rare (Lundy, 2013; Lundy, Crowe &

Turner, 2016; Simonton, 2011). Can the same things be said of beer critics? We applied a procedure previously used to study the music and film critics to this third area of aesthetics, that is, perceptions of beer quality.

Our archival study investigated appraisers of a wide variety of beers from all over the world, looking at over 700 beers in total. We predicted that, like previous studies of aesthetics in film and music, professional beer critics (experts) would tend to show positive correlations in their ratings (Hypothesis 1). Similarly, it was also predicted that beer critics would follow mound-shaped distribution patterns, approaching normality (Hypothesis 2). We also predicted that nonprofessionals (nonexperts) would in contrast show lower positive correlations that professional critics, and show less mound-shaped distributions (Hypothesis 3). Such findings have already occurred in a recent music study (Lundy & Smith, in press).

Method

We randomly selected over 700 beers from all existing beer critic books and websites that contained quantitative ratings. We then analyzed professionals' and nonprofessionals' ratings (N = 2,200 ratings). Professionals' ratings came from four books and one website (Campbell & Goldstein, 2010; Klein, 2000; Levinson, 1999; Robinson, 1996; BeerAdvocates Brothers on www.beeradvocate.com, 2017). Nonprofessional sources all came from websites, and were found on www.beeradvocate.com, www.ratebeer.com, and www.untappd.com. Potential quantitative markers of rating refinement were tested by analyzing frequency distributions of beer ratings. Hypothesis 1 will be compared to Lundy (2013) results for music. Hypothesis 2 will be compared to results for music (Lundy & Smith in press) and film critics (Lundy et al., 2016).

Results

Hypothesis 1 was partially supported. The majority of professionals showed significant positive correlations in their ratings (average $r = +.29$, 6 of 7 p's $< .05$), implying some agreement about beer quality, but overall agreement was lower for beer ratings compared to previous studies of music (+.49) and film (+.60) ratings by professionals (Lundy, 2013).

Hypothesis 2 was also partially supported. Critics' overall beer rating distribution was mound-shaped and approached normality, but it showed larger skewness (-.60) and larger kurtosis (-.37) compared to music (-.23 skewness and -.17 kurtosis; Lundy & Smith, in press), or film critics (-.40 skewness and +.02 kurtosis; Lundy et al., 2016).

Hypothesis 3 was partially supported as well. Professionals did show a rating distribution closer to normality, but were not higher in interrater reliability (r = +.27) compared to nonprofessionals (r = +.71)--in fact, the opposite was true. Overall, these findings extend research on questions of interrater reliability and rating distributions within aesthetics to a new aesthetic domain, namely beer quality appraisal.

Discussion

Similar to findings with music and film critics, we found that 1) beer critic pairs showed some consensus in their ratings, although this consensus was considerably lower than it was for music and film; and 2) that critics' overall beer rating distribution was mound-shaped and approached normality, although skewness and kurtosis were larger compared to music and film critics. The findings paralleled rating distribution characteristics wherein the majority of works were perceived by many raters to be moderate in quality, and greater and greater extremes of quality were perceived as less and less common. These findings do not fit with a strong "beauty is in the eye of beholder" stance, which would predict zero or even negative correlations of ratings.

Nonprofessionals surprisingly showed higher interrater reliability in comparison to professional critics. This may be because of nonindependent ratings: the nonprofessional ratings were all websites, whereas four of the five professional sources were books. There is the possibility of website consumers seeing and being influenced by each other's ratings, and is more likely for nonprofessionals. This would be a form of social influence or "conformity bias" (see Lundy, 2016).

Study Limitations and Future Research

One possible reason for lower correlations in this study could be that the beers compared were not identical because their exact nature could vary from year to year. Such a problem would be even greater if one attempted to do a parallel study of wine ratings. This issue is not the case in both the music and film industries wherein exact copies of works are available. There are still other areas of aesthetics to explore with respect to both agreement and the shape of rating distributions, such as visual art and literature, although quantitative ratings may be much more difficult to locate for these other areas of aesthetics.

References

BeerAdvocate.com (2017). *BeerAdvocate magazine*. Retrieved from: https://www.beeradvocate.com/

Blackburn, S. (1996). *Oxford dictionary of philosophy*. New York, NY: Oxford University Press.

Boor, M. (1990). Reliability of ratings of movies by professional movie critics. *Psychological Reports, 67*, 243-257. Retrieved from https://doi.org/10.2466/PR0.67.5.243-257

Campbell, S. & Goldstein, R. (2010). *The beer trials*. New York, NY: Fearless Critic Media.

Klein, B. (2000). *The beer lover's rating guide*. New York, NY: Workman Publishing Co.

Levinson, J. (1999). *Beer lover's companion: A guide to producing, brewing, tasting, rating, and drinking around the world*. New York, NY: Black Dog & Leventhal Publisher, Inc.

Lundy, D. E. (2010). A test of consensus in aesthetic evaluation among professional critics of modern music. *Empirical Studies of the Arts, 28*(2), 243-258. Retrieved from https://doi.org/10.2190/EM.28.2.h

Lundy, D. E. (2013). Critiquing the critics: Statistical analysis of music critics' rating distributions as a measure of individual refinement. *Empirical Studies of the Arts, 31*(1), 59-89. Retrieved from https://doi.org/10.2190/EM.31.1.d.

Lundy, D. E. (2016). Decontaminating taste: Minimizing nonaesthetic biases in aesthetic appraisal. *Review of Arts and Humanities, 5*, 1-16

Lundy, D. E., Crowe, A. C., & Turner, A. J., (2016). The shape of aesthetic quality: Professional film critics' rating distributions. *Psychology of Aesthetics, Creativity, and the Arts, 10*(1), 72-81. Retrieved from https://doi.org/10.1037/aca0000040

Lundy, D. E., & Smith, J. L. (in press). It's tough to be a critic: Professional versus nonprofessional music judgment. *Empirical Studies of the Arts, 35*(2), 139–168. Retrieved from https://doi.org/10.1177/0276237416661989

RateBeer LLC (2017). *RateBeer*. Retrieved from https://www.ratebeer.com/

Robinson, J. (1996). *The beer-taster's log: A world guide to more than 6,000 beers*. Pownal VT: Storey Publishing.

Simonton, D. K. (2011). Great flicks: Scientific studies of cinematic creativity and aesthetics. New York, NY: Oxford University Press. doi: http://dx.doi.org/10.1093/acprof:oso/9780199752034.001.0001

Untapped Inc. (2017). *Untapped*. Retrieved from https://untappd.com

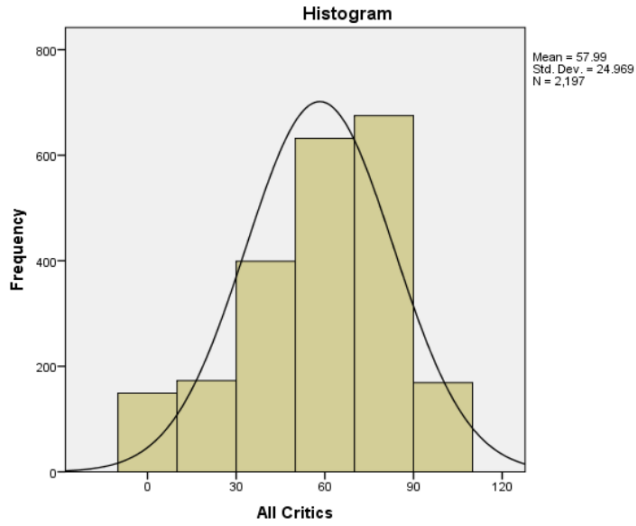

Figure 1: Normalized rating distribution of professional beer critics.

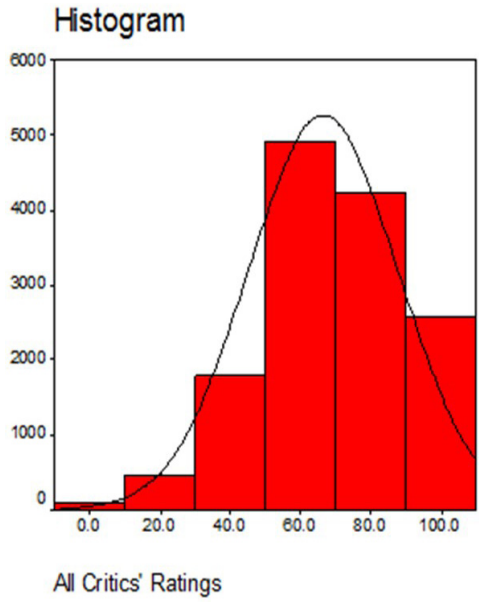

Figure 2: Normalized distribution example from previous aesthetics study from music critics (Lundy, 2013).

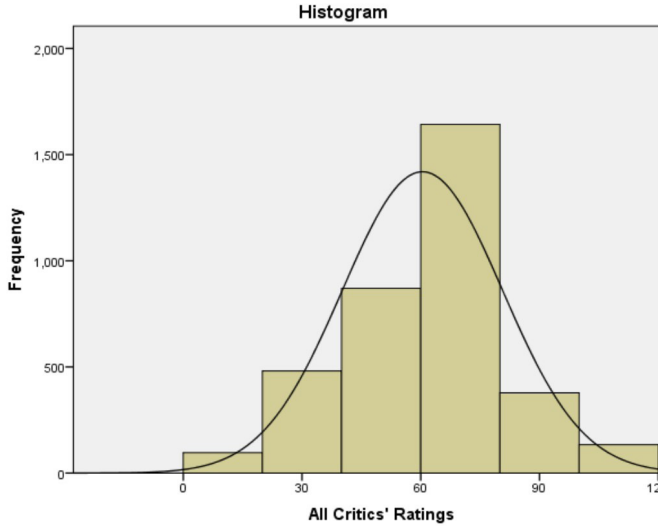

Figure 3: Another example of normalized distributions of ratings in aesthetics, from film critics (Lundy, Crowe, & Turner, 2016).

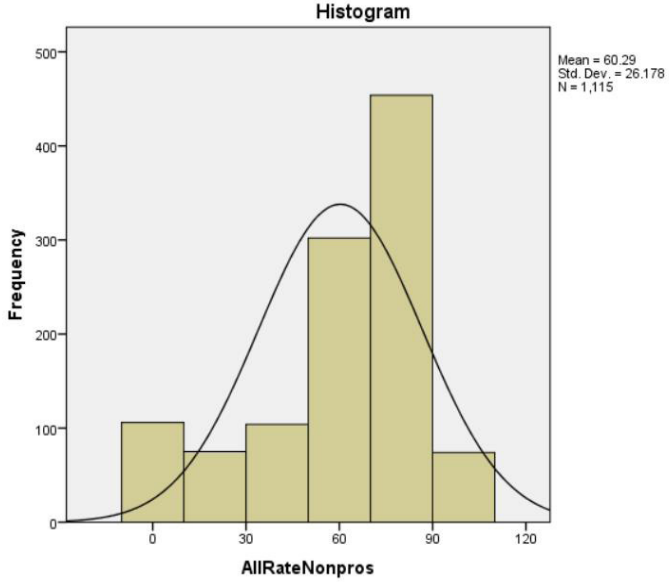

Figure 4: Nonprofessionals' rating distribution.

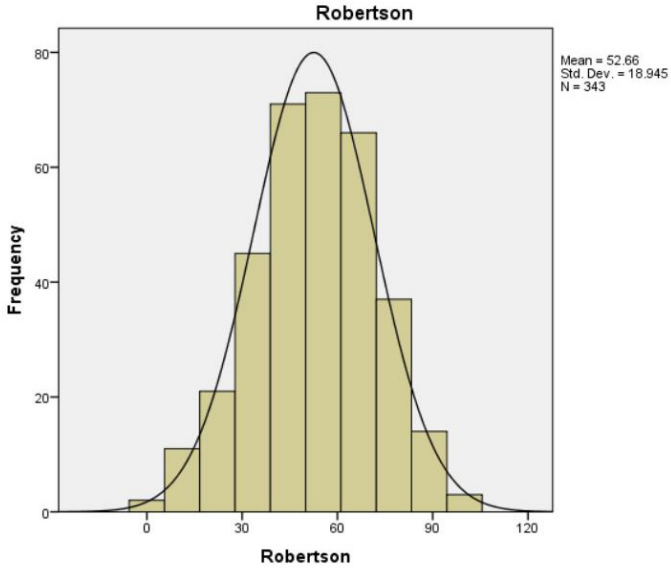

Figure 5: Normalized rating distribution of ratings from Robinson (1996).