Detecting Kantian Aesthetic Judgment in Film Reviews using NLP and Conditional Process Analysis

Björn Alvinge
Uppsala University
b.alvinge@gmail.com

Abstract

This paper is about testing a part of Kant's aesthetic theory, whether universally valid judgments of beauty are independent of personal preferences and desires. The importance of this is primarily the proof-of-concept, that a problem thought to only be within a philosophical-conceptual purview can be tackled using a data-driven approach. An NLP inference model is used to extract a set of variables from the text of film reviews, variables which statistical co-variance are then analysed with a regression model under a conditional process analysis framework. The findings indicate that interest as an effective cause for a judgment of beauty did not remain effective under a universal context. This vindicates Kant's aesthetic theory, under a certain interpretation. The results also purport to demonstrate the more general idea that for be a judgment to be valid for everyone, one needs to manage subjective attachment, something which professional film critics seem to do.

1 Introduction

> "In all judgements by which we describe anything as beautiful, we allow no one to be of another opinion." (Kant, 1987, p. 240)

Kant's statement here might seem puzzling to contemporary readers. Is it not true that beauty is in the eye of the beholder? While Kant agreed that judging something as beautiful entails relying upon subjective feelings, like taking pleasure in an object, Kant's statement above serves to point out a that such a pleasurable experience in the beautiful has universal properties. That is, in the statement "x is beautiful", there is a tacit demand that everyone can share in that experience by reflecting upon the object's form. As such, x in this case is not something for a select few who harbour a certain set of private wants and desires, which is what informing their personal opinions about what is beautiful to them. Beauty is a pleasurable experience which show universal communicability, and thus a demand for universal assent for it is possible (Ginsborg, 2022). Kant wanted to point this out, so as to ward off against misuses of the term "beauty" which did not meet this universal criterion (Daniels, 2008). Kant saw that a common "misuse" of the word "beauty", due to its strong subjectively pleasurable experience, could be found in that people would come to mistakenly believe that something could be regarded as "beautiful" if it was pleasurable. This happened because people used a subjective criterion for judging something a beautiful, namely a sensuous pleasure that was of their personal preferences, related to their desires and inclinations being met in an object. Such private considerations for beauty could not transcend the level of personal opinion. This subjective criterion for beauty, was an "interested" form of aesthetic judgment. His conclusion thus became that, in order to make sure one was open to beauty which retained its universal properties, one needed to make sure one was not affected by one's "interests" in one's aesthetic judgment. One had to possess a level of "disinterestedness" in one's aesthetic judgment for it to be of universal validity.

This particular sense of "disinterestedness" as a concept is synonymous with being impartial and objective.[1] The concept is also synonymous with a type of detachment in one's aesthetic judgment. These explications links up with Kant's, in that not showing interest in one's aesthetic judgment, is a what it means to be less bound by a subjective criterion for positive judgment, and thus being more objective. In Kant's terms, only through

[1] see https://www.merriam-webster.com/dictionary/disinterestedness

disinterestedness can a universally valid aesthetic judgment be generated, and it is pleasure in the beautiful, that is disinterested (Ginsborg, 2022).

The research question of this paper is to see if it possible to find empirical evidence of this thesis that a universally available or shareable judgment of beauty is disinterested, by using a set of modern techniques from language technology and statistics applied to film reviews. The test involves analysing to what extent interest affect a judgment of beauty under the quasi-experimental condition of an increasing level of universal context. The research hypothesis is that this effect should dissipate under a universal context, according to the theory that an interested judgment of beauty is empirically contradictory with it being universal, which is thought to provide evidence for the thesis that universally valid judgments of beauty are disinterested.

1.1 Overview

The experimental test went through 3 phases. First, the level of interest, beauty, universal communicability and other variables being expressed in a set of film reviews were measured and coded as probability variables. Second, interest was included as the primary explanatory variable, and the others as controls/moderators, with beauty as the dependent variable, in a statistical model. Third, it was measured if under a conditional process analysis (CPA) with that statistical model, with universal communicability and meta score as moderators, if interest was always effective in explaining the outcome of beauty in a review. The result show that the effect disappears as universal communicability and meta score gets to a certain level. However, to what theoretical extent the the analysis show empirical evidence for Kant's thesis, is discussed.

2 Related Work

The issue of disinterested and universal aesthetic judgment has mostly been analysed with a philosophical/conceptual method (Ginsborg, 2022). However, a neurological experiment have empirically verified an aspect of Kant's aesthetic theory relevant to this paper (Brielmann and Pelli, 2017). They demonstrated that order to appreciate beauty, one needs to engage in conscious thought. By distracting participants in second phase with a memorisation task as aesthetic judgments were being made, participants felt less pleasure viewing beautiful images, and indifferent about plain one's. This corroborates the picture how Kant thought beauty was properly appreciated, i.e. not by simply consulting one's personal, subjective sense pleasures for the object, but primarily through a conscious reflection upon the object's form. Kant argued that the universal communicability of beauty was possible due to a reflection upon the object's form (Ginsborg, 2022). Because Kant thought we universally share this reflective cognition, we can universally communicate that object's form to others so as to make others unequivocally share in our aesthetic pleasure, but only as long as there was no interest in that pleasure, as then that pleasure would be a private matter, and not be universally communicable. The NLP aspect of this study is equivalent to classic sentiment analysis of film reviews. Sentiment analysis involves extracting the sentimental polarity of a text, i.e. whether a text is expressing a positive or negative sentiment. There exists a wide array of methods and procedures for extraction (Lakshmi Devi et al., 2020), and using specifically large language model to infer a theoretically nuanced definition of sentiment can also be found (Clarke et al., 2023). Work that uses the same proposed analytical method to solve a different problem can be found in the the statistical aspect of this study. Conditional Process Analysis (CPA) is an analytical strategy in order to test hypotheses about how mechanisms vary as a function of context or individual differences (Hayes and Rockwood, 2020). For example, (Palmer et al., 2016) show that the direct effect that recalled academic experience has upon loyalty to a university as a brand decrease as a function of time. In this study, we are interested in a direct effect, and how this changes as a function of a universal context. predicting a continuous value from text using a regression model is available (Cohen, 2022). The CPA in this study uses a regression model (Hayes, 2017). Combining numerical and text information, which is done in this study, can be found in order to make causal inferences (Ahrens et al., 2021). However, a method that is similar to this method that solves a relatively similar research problem has not been found. There is no example that analyses film reviews with CPA to our knowledge, and especially not to tackle a problem strictly from the domain of philosophy, treating it

as an empirical problem to be analysed using film reviews.

3 Theory

The research hypothesis can be broken down into two testable arguments for their case:

3.1 Null-Hypothesis: The Strong Version of Subjective Beauty

H0: That which one person finds beautiful may not appeal to another. That which one person finds beautiful, is necessarily appealing, only to them. What is appealing to someone means having personal preferences and desires for something, i.e. showing "interest". Therefore, what one person finds beautiful, is always a function of their interest. The conclusion can also be stated as what is beautiful is always a matter of personal opinion. Universality must under such a rubric mean that a personal preference is still influencing the judgment of beauty, even when it is universally shareable.[2]

H0 argues for a strong version of beauty being in the eye of the beholder, that it has a specific subjective criterion i.e. interest, and this subjectivity is hypothesised to not be empirically contradictory with a universal context.

3.2 Alternative-Hypothesis: The Weak Version of Subjective Beauty (Objective Beauty)

H1: That which one person finds beautiful may not appeal to another. But that which one person finds beautiful, is not due to that person finding the object appealing to them, as a type of detachment is necessary to appreciate beauty's truly universal character. What is appealing to someone means having personal preferences and desires for something, i.e. showing "interest". A universal judgment of beauty cannot co-vary with interest, because a personal preference as a criterion for a judgment of beauty would make that beauty a mere personal opinion, which cannot be available or applicable to everyone.[3] Therefore, what one person finds beautiful, is not always a function of their interest. What is (universally) beautiful, is not a matter of personal opinion.

H1 argues for a weak version of beauty being in the eye of the beholder, in that it involves subjective, pleasurable feelings to an extent, but interest as a subjective criterion for beauty is hypothesised to be empirically contradictory with a universal context.

H0 argues for a judgment of beauty being universally valid if it is informed by inter-subjective personal preferences of universal scope, whilst H1 argues for the source of universally valid judgment of beauty to be a preference-independent, reflective objectivity. H1 argues for the notion that in order to retain the aesthetic judgment's validity for everyone, a type of subjective detachment is required, whilst H0 does not think such detachment is neither necessary nor possible.

4 Method

4.1 Raw Dataset

The raw dataset for the movie reviews can be found on Kaggle.[4] It contains at least 12029 movie titles (some movies have the same title, so there is slightly more films present), with a total of 233198 reviews. Each review has a column for individual_meta_score, which is the evaluation score from 0 to 100 given by a critic, the critic_name, movie_title, and the actual film review text. This data set was primarily chosen because it contains aesthetic reviews made by professional critics, a necessary ingredient in measuring the right kind of variables that will be used in a regression model.

4.2 Measurement of Text Variables

A natural language inference model was used to measure the text-extracted variables in this study.[5] This model is a large Transformer model, having been pre-trained on a multi-genre dataset with the task of inferring whether a hypothesis string contradicts, is neutral to, or is entailed by a premise string. After having been pre-trained, the model

[2] For example, humans find symmetry almost universally appealing when judging if someone is beautiful as in attractive (Grammer and Thornhill, 1994). Beauty in the sense of attractiveness is a subjective criterion of "beauty". It thus the case that a specific subjective sense of "beauty" is universally shared, shared as a hard-wired preference in everyone. But whether it is possible as a matter of aesthetic practice to make judgments of beauty which are informed by personal preferences based on desire, and at the same time have them be universally communicable, remains to be seen.

[3] Since beauty is thought to be universally available due to a reflection upon the object's form for Kant (Ginsborg, 2022), and not through agreement of sense, a personal preference should not have explanatory power in predicting the type of beauty that is universally available.

[4] https://www.kaggle.com/datasets/miazhx/metacritic-movie-reviews

[5] https://huggingface.co/facebook/bart-large-mnli

outputs a probability measure from 0 to 1 depending on whether the hypothesis string, or the prompt, is entailed in the premise string, the text. The pre-trained model was used in a zero-shot classification/inference scheme, used for the task of measuring a set of independent variables and a dependent variable, from the movie review text. There are in total 6 probability variables constructed from the text, 5 of which are used in the regression model. Those 5 are constructed by using "This text contains x" as prompt, where x is replaced with the variable names:

1."beauty"(B): Whether the text talks about a judgment of beauty, or taking pleasure in the beautiful.

2."displeasure"(D): Whether the text talks about a displeasure in the film.

3."negativity in sentiment"(NIS): Whether the text talks about having a negative sentiment or opinion about the film.

4."A sheer personal preference based on a desire or inclination"(I)[6]: Whether there is an interest in the film by the critic, based on desire. Similar to a liking for the film or whether she is "charmed" by it. This variable name is replaced with "interest" as shorthand after the probability scores has been measured.

5."universal communicability"(UC): Whether the text talks about an experience communicable, transferable, shareable, available by or for everyone.

This effectively produces a 0 to 1 probability score for each variable name, i.e. whether the premise or review text contain any of the variables. Disinterestedness itself is a negative causal concept, and is assumed to be about the interaction of the interest variable with the beauty variable. Disinterestedness is thus tested with the forthcoming regression model. The inference model was chosen primarily because of the simple extraction process of the variables, along with the easy interpretability and statistical usefulness of the probability measures produced. The final data set used in the statistical model, with each variable as columns, can be seen in Table 1.

4.3 Statistical model

An OLS multiple regression model, model nr. 2 from the Python implementation[7] of Hayes Process Macro (Hayes, 2017), is used for the conditional process analysis. This model employs one x and one y variable, two non-interactive moderators, along with controls. The model's equation is:

$$Y = b0 + b1X + b2W + b3Z + b4XW + b5XZ + b6C1 + b7C2 \quad (1)$$

Interest(I) is the main explanatory/independent variable X and beauty(B) is the dependent/explained variable Y. Universal Communicability(UC) and meta-score are the moderators W and Z. The controls are Negativity In Sentiment(NIS) and Displeasure(D) are C1 and C2.

4.3.1 Explanation of Model

The statistical model is set up so as to test in particular causal direction, namely in the subjectivist/"misused" direction of beauty. We are interested to see if H0 is false under a certain condition, in that interest should not exert an effect upon a judgment of beauty, i.e. when beauty has interest as a subjective criterion, under a universal condition. H1 Assumes that UC, and the meta-score, are variables that moderate the directional relationship between interest and beauty, in such a way that interest should not be able to explain/predict/affect the type of beauty which is universal. UC codes for if an experience can be universally shared, but meta score probably also codes for universality.[8] Finally, the variables D and NIS are used as control variables. These are used, as it was found that a naked I produced a negative effect on B. This happened because I correlated somewhat with the control variables, in that a personal preference would a have high probability when the sentiment was negative. This is probably reflecting the fact that a negative or displeasureable experience is still itself coding for a personal preference (see Table 2). In this study however, we are primary concerned with the positive effect as a positive subjective criterion, that I should, or

[6]This formulation is taken from this YouTube video: https://youtu.be/6lOwHdydgi4?t=70

[7][Link]

[8]If we browse the top scoring movies here: https://www.metacritic.com/browse/movie/, almost everyone of them is universally acclaimed above a certain average meta-score. It can be argued that this universal acclaim is potentially carried on an individual level, when the critic gives a high score. A possibly simpler explanation is that a high meta score co-varied with high UC (when UC=0.9 and above is selected, meta-score is (M = 75.55, SD = 15.23)). As such, they in concert code for the same type of universality.

meta score	text	B	I	UC	D	NIS	AFNOTD
100	It's one of the year's most galvanizing experiences.	0.9696	0.0096	0.8654	0.0474	0.0010	0.9285
100	Utterly magnificent and intoxicating.	0.9973	0.8192	0.9915	0.0045	0.0007	0.9340
67	This hankie-yanker is an emotional cheat.	0.0075	0.9919	0.0030	0.9947	0.9937	0.0027
80	A poetic, though admittedly esoteric piece of cinema.	0.9994	0.6257	0.0093	0.0023	0.00090	0.0035
74	When Team America works, it falls squarely into the category of guilty pleasure.	0.5900	0.9935	0.5304	0.0235	0.0204	0.2271
88	Clever, buoyant and surprisingly human.	0.9994	0.6382	0.9959	0.0185	0.0007	0.0002

Table 1: An excerpt of the data set with the variables used in the statistical model, except for "A statement which allows for no one to disagree"(AFNOTD), which is first mentioned in the result section.

should not, exert on B. We thus adjust for them, so as to remove their negative effect.

4.3.2 Conditional Process Analysis

Using the previously mentioned regression model, a Conditional process analysis is set-up. The mechanism in this study is the effect interest has upon beauty, and the context is under a universal (UC and meta-score) context. We are thus interested in the conditional direct effect that X exerts on Y under each value of the moderators W and Z, i.e. how X changes Y depending upon W and Z. Conditional Process Analysis allows us to do this. Some algebra[9] to show this:

$$Y = b0 + b1X + b2W + b3Z + b4XW + b5XZ + b6C1 + b7C2 \quad (2)$$

Grouping terms into form Y = a + bX:

$$Y = (b0 + b2W + b3Z + b6C1 + b7C2) + (b1 + b4W + b5Z)X \quad (3)$$

Direct effect of X on Y, conditional on W and Z:

$$b1 + b4W + b5Z \quad (4)$$

4.3.3 Explanation of Equation(s)

To get the conditional direct effects, we first look at the OLS regression model to get the beta values. These are the coefficients, i.e. average values of how the dependent variable changes for each independent variable as others are held constant. We also acquire the VCV values from the model, which is the variance-covariance matrix which is used to calculate the standard error for the variables. Then, the partial derivative of the equation in (3) with respect to X i.e. Interest(I) is computed, which is what is seen in (4). This equation is then evaluated at the different values of the moderators. Then, the dot product of those gradients/evaluations and the beta values are calculated to get the conditional direct effect. VCV is also computed with in a similar way by computing the dot product of VCV and the gradients to get the standard error at the different values of the moderators, which is then calculated together with the conditional direct effects mentioned earlier in order to get the p-values and thus what the significance is for those effects. In process macro, the result from the conditional process analysis is an output which contain the conditional direct effect that X has upon Y under different values of the moderators, along with the p-values for the significance of that effect. [10]

5 Results

At high values of Universal Communicability (UC) and meta-score, the direct effect that interest has upon beauty, along with significance of the relationship, disappears, as can be seen in Table 2. Furthermore, an observation after this analysis was made showed that, if the prompt "A statement which allows for no one to disagree"(AFNOTD)

[9] http://www.figureitout.org.uk/model2.htm

[10] This explanation was made by reading the source code for PyProcessMacro, found in footnote 7.

i.e. universal assent was measured using the inference model, and if data with above 0.9 for this variable was selected, this data showed a high meta score (M = 75.32, SD = 23.54) along with a somewhat high UC (M = 0.572, SD = 0.395).

6 Discussion

The purpose of this study was to see if disinterestedness could be detected in a judgment beauty under a universal context using modern NLP and statistical tools. The statistical/analytical results indicates such a detection, in that a judgment of beauty cannot be affected by interest, a personal preference based on desire, and at the same time be universally communicable. This casts doubt upon the null hypothesis, in that a judgment of beauty is a matter of personal preference, and vindicates the alternative hypothesis, in that when a beauty is judged to be for everyone, this universal quality can only be rendered when that judgment is independent of, or detached from, the critic's personal preferences based on his desires. The observation indicates that the different forms of universality, assent, communicability and acclaim, perhaps overlap. The observation and analysis thus provide evidence for the case that a judgment of a universally shareable beauty is not a matter of sharing a certain personal opinion or preference for an object, and is not something one can disagree with as a matter of personal taste. However, there are questions that can be raised for such a bold set of assertions being made via an admittedly peculiar method. First, it is important to stress that the research problem does not exhaust the entirety of Kant's aesthetic theory, as it only tests the beginning moments of it, i.e. that Kant said that only from a disinterested judgment can a universal judgment follow (Ginsborg, 2022). Second, one can also discuss whether the method of this paper demands a comparison with another dataset, for example with non-professional reviewers. However, the research question is not about whether professional critics do something different compared to non-professionals, but whether aesthetic judgments in general, which professional critics make, have a particular form or not under a universal context. One could argue there already is a group comparison being made here, where each group can be found at different levels of the meta-score and universal communicability. But perhaps there is something illuminating about such a professional v.s. non-professional group comparison in the final analysis. Third, Table 1 provide some ideas of how the variables are interpreted, but do not entirely exhaust all of the examples that the NLI model gave a high probability to. Currently, it is only the CPA and the curated observations that show that the right theoretical interpretations are being made by the NLI model, as a matter of how the variables seem to interact. But one can argue that is not sufficiently strong argument for not making a more sophisticated subjective evaluation of the variables in isolation. Furthermore, AFNOTD should probably have had it's own regression analysis in order to check for proper association with the relevant variables. Finally, one can also discuss the choice of model structure and the theory in this paper. Model structure is informed by this paper's theory, which is a slightly modified version of Kant's theory, following the explanation by Daniels (2008). Kant never explicitly spoke of subjective criteria for beauty, although this is the principled issue Daniels think Kant was worried about, a version of beauty which did not render any universal qualities. Future research would have to look at better ways to ensure valid capture of all variables used, in order to get closer to the notion that universally valid aesthetic judgments are disinterested. Perhaps also a stronger causal inference or discovery framework needs to be established, considering the implied causal directionality of the theory and the model.

7 Conclusion

This paper was about testing whether the first moments of a slightly modified version of Kant's theory was correct, using a data-driven and statistical approach upon film reviews. It was found that a subjective criterion for beauty, interest, lost its direct effectiveness as conditional contexts which coded for universality became more prevalent. These results empirically vindicate Kant's theory that aesthetic judgments are disinterested; That a detachment from personal preferences and desire in one's judgment of beauty is necessary for that judgment to retain it's universally valid qualities. However, further research would have to look into validity of the measurement of the variables, together with the theoretical leaps of interpretation of the statistical and analytical results, in order to corroborate these empirical findings.

Spotlight Moderator (meta-score)	Focal Moderator's Moment of Insignificance(UC)	Conditional Direct Effect At UC=1
83	0.9929	0.0109
84	0.9822	0.0105
85	0.9712	0.0101
86	0.96	0.0097
87	0.9484	0.0093
88	0.9367	0.0089
89	0.9246	0.0085
90	0.9122	0.0081
91	0.8996	0.0077
92	0.8866	0.0073
93	0.8734	0.0069
94	0.8598	0.0065
95	0.8459	0.0061
96	0.8317	0.0057
97	0.8172	0.0053
98	0.8023	0.0049
99	0.7871	0.0045
100	0.7715	0.0041

Table 2: This table show under what moderator values the direct effect interest(I) has upon Beauty(B) become statistically insignificant (alpha=0.05). Between 0-82 in meta score, the effect remain significantly positive for any value of UC (7.106e-05 to 0.9998), but at 83 and at high probability of UC (0.9929 to 1), the effect start to disappear. This process then continues, as the insignificance interval starts to go lower and lower in scope for UC as meta-score gets higher, up to 100 in meta score which is the highest score that can be given. The subjective effect's disappearance as a function of higher meta-score can be seen in the 3rd column, specifically when UC=1 at the same moment.

References

Maximilian Ahrens, Julian Ashwin, Jan-Peter Calliess, and Vu Nguyen. 2021. Bayesian topic regression for causal inference. *arXiv preprint arXiv:2109.05317*.

Aenne A Brielmann and Denis G Pelli. 2017. Beauty requires thought. *Current Biology*, 27(10):1506–1513.

Patrick Clarke, Carly Leininger, Cristiana Principato, Patrick Staples, Guy M Goodwin, Gregory A Ryslik, and Robert F Dougherty. 2023. From a large language model to three-dimensional sentiment.

Shay Cohen. 2022. *Bayesian analysis in natural language processing*. Springer Nature.

Paul Daniels. 2008. Kant on the beautiful: The interest in disinterestedness. *Colloquy*, (16):198–209.

Hannah Ginsborg. 2022. Kant's Aesthetics and Teleology. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*, Fall 2022 edition. Metaphysics Research Lab, Stanford University.

Karl Grammer and Randy Thornhill. 1994. Human (homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. *Journal of comparative psychology*, 108(3):233.

Andrew F Hayes. 2017. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Andrew F Hayes and Nicholas J Rockwood. 2020. Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist*, 64(1):19–54.

Immanuel Kant. 1987. *Critique of judgment*. Hackett Publishing.

B Lakshmi Devi, V Varaswathi Bai, Somula Ramasubbareddy, and K Govinda. 2020. Sentiment analysis on movie reviews. In *Emerging Research in Data Engineering Systems and Computer Communications*, pages 321–328. Springer.

Adrian Palmer, Nicole Koenig-Lewis, and Yousra Asaad. 2016. Brand identification in higher education: A conditional process analysis. *Journal of Business Research*, 69(8):3033–3040.