

Quantitatively evaluating changes in art critique journals' use of moral language across time

Emerson Chang¹, Marwa Mahmoud²

¹ Henry M. Gunn High School, Palo Alto, California

² University of Glasgow, Glasgow, Scotland, UK

SUMMARY

In art journals, art critics examine the connection between the arts and their social contexts. Although sentiment analysis techniques have been applied to articles in this domain, a quantitative evaluation of their moral language has yet to be conducted. We collected approximately 5000 articles from seven art journals published between 1960 and 2015. Relying on the Moral Foundations Theory (MFT) and the Moral Bidirectional Encoder Representations from Transformers (MoralBERT) model, we evaluated trends in the art journals' moral language throughout this period. The MFT outlines the psychological systems that allow humans to successfully survive and reproduce, and it proposes the existence of "Virtues" and "Vices." "Virtues" help individuals navigate social relationships and hierarchies, while "Vices" have the opposite effect. MoralBERT outputs values from 0 to 1 for each of 10 "Virtues" and "Vices." A higher value indicates that the text expresses a foundation to a greater extent. We hypothesized that the average foundation probabilities of the "Virtues" would decrease, while the foundation probabilities of the "Vices" would increase. The results revealed that, although changes across time were limited, the frequency of the "Purity" foundation, which aligns with a desire for sacred cleanliness, decreased. Meanwhile, the "Degradation" and "Care" foundation probabilities increased across time. "Degradation" entails contamination, in both a physical and spiritual sense. The "Care" foundation motivates individuals to alleviate others' pain through acts of kindness. Overall, this study reveals how artists and art critics' priorities have shifted across time.

INTRODUCTION

Art journals provide scholarship and exploration in the visual arts, highlighting the perspectives of artists and art historians (1). Art journals respond to major turning points in the realm of art, and they illustrate modern issues by placing works of art in the broader social context (1). Art itself is also a tool for communicating with audiences. For example, when viewers saw art about climate change, they demonstrated more support for climate policies in the short run (2). Specific features of the artwork impacted first impressions, personal reflections, and the strength and type of emotions elicited (2).

Similarly, art journals can convey sentiment; positive and negative sentiments align with positive and negative emotions, respectively. Sentiment analysis studies evaluated art articles that were published after 1960 because the period corresponded with an increase in social movements, including human rights, women's rights, and gay rights (3). Art is often

created in reaction to social developments, and art critics analyze artwork in the context of these events. From 1960 to 2015, the emotions embedded in art review titles trended in a positive direction, expressing supportive language toward the art being discussed (3). Using TextRank, Bao and Gao generated three-paragraph summaries of the articles, and these three paragraphs constituted the abstract of the articles (3). Between 1960 and 2015, articles' sentences and abstracts expressed less positive sentiment than the articles published at the beginning of the period (3). For example, in the full texts, the average sentiment decreased from 0.30 to 0.24 across the 55-years.

Sentiment analysis is a popular topic in computer science. It entails detecting writers' opinions on specific topics (4). First, a corpus of documents is rendered as text. During the subsequent preprocessing stage, computational models derive root words, divide the text into tokens, assign grammatical categories to words, classify entities, and examine relations between entities (4). Next, sentiment annotations are attached to the whole document, individual sentences, or specific aspects of the text (4). Sentiment analysis can also be applied to social media posts. For example, Nemes and Kiss used a Recurrent Neural Network to identify emotions in tweets related to the COVID-19 pandemic (5). They analyzed emotions along a spectrum with positive/negative, weakly positive/negative, strongly positive/negative, and neutral values (5). Positive tweets offered support and hope to those in distress, while negative tweets discussed the deaths resulting from the pandemic (5).

The Moral Foundations Theory (MFT) outlines the universal moral foundations that individuals from all societal and cultural backgrounds possess from birth (6). Over time, individuals have been exposed to repeated problems and opportunities, leading them to develop functionally specialized mechanisms for quick and effective responses (6). The moral foundations underlie a rudimentary form of morality and enable individuals to solve social problems (6). Organized in "Virtue" and "Vice" pairs, the original MFT from 2013 juxtaposes "Care" with "Harm"; "Fairness" with "Cheating"; "Loyalty" with "Betrayal"; "Authority" with "Subversion"; and "Sanctity" (also known as "Purity") with "Degradation" (6). "Care" entails sympathy, compassion, and nurturing directed toward those in distress (6). "Harm," meanwhile, is a violation of "Care," prompting anger toward the perpetrators (6). "Fairness" reflects the desire for equal transactions between partners (6). "Loyalty" encompasses the formation of cohesive coalitions for a tactical advantage (6). "Authority" corresponds with the ability to navigate dominance hierarchies and form advantageous relationships; individuals also express obedience toward those in power (6). Finally, "Sanctity" results in feelings of disgust, aversion to illness, and spiritual purity (6). In addition

to these five pairs proposed in the original MFT, “Liberty/Oppression” has been proposed as a potential foundation pair (6). “Liberty/Oppression” leads individuals to express anger toward powerful oppressors (6).

Moral judgment enables the appreciation and evaluation of art, and art contains moral content (7). In writing and speech, individuals emphasize moral foundations that align with their values. For example, liberal speakers promoted “Harm” and “Fairness” to a greater extent than conservative speakers while giving church sermons (8). These differences were not specific to writing, as liberals’ and conservatives’ values differed when they used numbered scales to rate the relevance of different moral considerations, indicate their agreement with moral judgment items, and report their unwillingness to perform actions that violated “Virtues” (8).

The detection of moral foundations has spread to a variety of domains, including politics. Clifford and Jerit investigated moral language in *New York Times* articles discussing stem cell research (9). They annotated the articles for words that aligned with “Harm,” “Purity,” and general moral language derived from the Moral Foundations Dictionary (MFD) (9). Clifford and Jerit used the term “Harm” to refer to both the “Care” and “Harm” foundations (9). They determined whether words aligned with evidence for or against stem cell research, as well as whether the writer upheld or rejected the evidence (9). Those who supported stem cell research tended to express only the “Care/Harm” foundation (9). Meanwhile, opponents conveyed “Care/Harm” language, general moral language, and some “Purity” language (9). Supporters and opponents used “Care/Harm” words in different contexts (9). For example, supporters invoked “Care” while discussing the advantages of saving lives, alleviating pain, and having compassion toward patients who could benefit from stem-cell therapy (9). Opponents, on the other hand, used “Harm” (violation of “Care”) to argue against destroying, exploiting, and killing embryos (9). During an interview from a prior study, respondents indicated whether they prioritized the medical cures that could result from stem cell research or if they would rather protect embryos by avoiding the research (9). Clifford and Jerit counted the number of moral words supporting and opposing stem cell research in *New York Times* articles published a week before each person’s interview (9). Public support for both sides of the argument increased following these debates, sparking interest in people who were previously uninformed or neutral about stem cells (9). As the number of Total Proponent Moral Words (moral words used in support of stem cell research) in the news articles increased from 0 to 10, the probability of respondents favoring medical cures increased from 67% to 77% (9). Meanwhile, as the number of Total Opponent Moral Words (moral words used to argue against stem cell research) increased from 0 to 10, the probability of respondents favoring medical cures decreased from 67% to 62% (9). Nevertheless, respondents’ answers could have reflected their confirmation biases—when no moral language existed in the news articles, they already displayed a 67% probability of prioritizing medical cures (9). Those with pre-existing moral beliefs tended to disregard opposing views (9). Clifford and Jerit’s findings demonstrate how individuals write arguments in response to moral issues and support their perspectives with moral language. Therefore, studying art critics’ articles allows us to understand their reactions to the moral narratives embedded in art.

Different approaches have emerged for extracting moral language from text. The MFD consists of 295 words selected by the authors of the MFT (8). Consulting thesauruses and colleagues, they identified foundation-supporting and foundation-violating words, along with their synonyms, antonyms, word stems, and deviational morphemes (8). Later, the researchers eliminated words that held non-moral double meanings (8). The MFD follows a bag-of-words approach and assigns one foundation to each word (10). Meanwhile, the extended Moral Foundations Dictionary (eMFD) incorporates crowd-sourced human annotations of news articles, which were highlighted for moral foundations (10). The eMFD thus allows words to appear in context (10). One word can belong to multiple foundations (10). Each word in the eMFD contains foundation probabilities and sentiment scores (10).

Bidirectional Encoder Representations from Transformers (BERT) is a model that can convert human language into vector representations (11). BERT randomly masks input tokens and predicts a token from its context, both preceding and succeeding (11). By analyzing the complete context around tokens, BERT can understand language to great accuracy and depth (11). BERT can be used for sentiment analysis and named entity recognition (11). BERT has also been incorporated into the domain of moral language with the creation of MoralBERT (12). MoralBERT predicts moral foundation probabilities from text (12).

Building on previous studies, we used MoralBERT to evaluate trends in the moral foundations expressed in seven art journals across a period of 55 years. The current MoralBERT model does not allow for the evaluation of “Liberty/Oppression,” and this foundation pair was thus excluded (12). This is the first study to quantitatively evaluate if and how moral values have shifted across time in art journal articles. Throughout the course of human evolution, moral “Virtues” have provided advantages to individuals, and the general expression of “Virtues” may have increased across time (6). Nevertheless, a previous study indicated that the average sentiment of full-text art journal articles decreased from 1960 to 2015, becoming less positive (3). Therefore, we hypothesized that the average foundation probabilities of each of the “Virtues” would decrease from 1960 to 2015, while the average foundation probabilities of each of the “Vices” would increase across this period. The years between 1960 and 2015 constitute a relatively short time period in the course of human evolution and could reflect a temporary decrease in “Virtues” relative to an overall increase across history. Additionally, a small set of individuals produced the art criticism texts, so these texts do not reflect the moral values of human society as a whole.

We found a decrease in the “Purity” foundation probabilities and an increase in the “Degradation” and “Care” foundation probabilities across time. By applying a pre-trained machine learning model to a new domain, we measured the effectiveness of the model in identifying the art journal articles’ moral values. Overall, our study holds importance for understanding the moral values that art critics emphasize. As everyday art enthusiasts examine works of art, they can apply the Moral Foundations framework to gather insights into the pieces. They can also reinterpret the works by considering which “Virtue” or “Vice” an art critic would have focused on. These new perspectives allow for a more enriching experience while viewing art.

RESULTS

MoralBERT provided “Virtue” and “Vice” foundation probabilities for each article, considered as a whole (12). The “Virtues” are “Care,” “Fairness,” “Loyalty,” “Authority,” and “Purity” (12). The “Vices” are “Harm,” “Cheating,” “Betrayal,” “Subversion,” and “Degradation” (12). The probability of “Degradation” occurring in the texts remained higher than the other four “Vices,” as “Degradation” trended in a slight positive direction, rising from 9.07% in 1960 to 12.74% in 2015 (**Figure 1**). The overall probabilities of moral “Vices” appearing in the texts remained low, with the highest probability being 13.28% for “Degradation” in 2005 (**Figure 1**). The low probabilities indicated a lack of moral language. In general, changes across time were also minimal for “Harm,” “Cheating,” “Betrayal,” and “Subversion.”

Similar to the “Vices,” the probability of the “Virtues” appearing in art journal reviews remained low, with no foundation probability rising above 7% (**Figure 2**). The low probabilities of the “Virtue” foundations indicated that the art journal articles contained limited moral language. From 1960 to 2015, the level of “Care” expressed in art journals increased to a small extent, with the “Care” probability rising from 5.21% in 1960 to 6.78% in 2015 (**Figure 2**). On the other hand, the “Purity” foundation illustrated a downward trend, while “Authority” and “Fairness” fluctuated upward and downward, albeit within a small percentage range (**Figure 2**). “Loyalty” remained roughly constant (**Figure 2**). The use of

moral “Virtues” exhibited little change over time.

We created scatterplots with linear trendlines to analyze slopes and determine trends in foundation probabilities over time. Positive and negative slope signs suggested upward and downward trends, respectively. “Harm” ($p < 0.05$), “Betrayal” ($p < 0.05$), and “Degradation” ($p < 0.01$) had positive slopes, indicating that their presence in art journals increased over time (**Figure 3**). “Subversion” and “Cheating” were not correlated with time (**Figure 3**). While the “Purity” foundation’s slope was negative ($p < 0.001$), the “Care” ($p < 0.001$) and “Loyalty” ($p < 0.05$) foundations had positive slopes of probabilities against time, suggesting that the expression of “Care” and “Loyalty” generally increased in art journal articles from 1960 to 2015 (**Figure 4**). The other “Virtues,” “Authority” and “Fairness,” had nonsignificant relationships (**Figure 4**). For both the “Virtues” and the “Vices,” the slope values were small due to the limited range of probabilities (0 to 1) compared to the wide range of years (1960 to 2015).

We evaluated the most frequent score ranges for each moral foundation. The “Care” and “Degradation” foundations’ most frequent score ranges were higher than those of the other foundations (**Table 1**). It was more likely to find “Care” and “Degradation” scores in this higher range than in the lower range exhibited by other foundations. Thus, “Care” and “Degradation” appeared more frequently in the texts than the other foundations.

When comparing the foundation scores for each of the

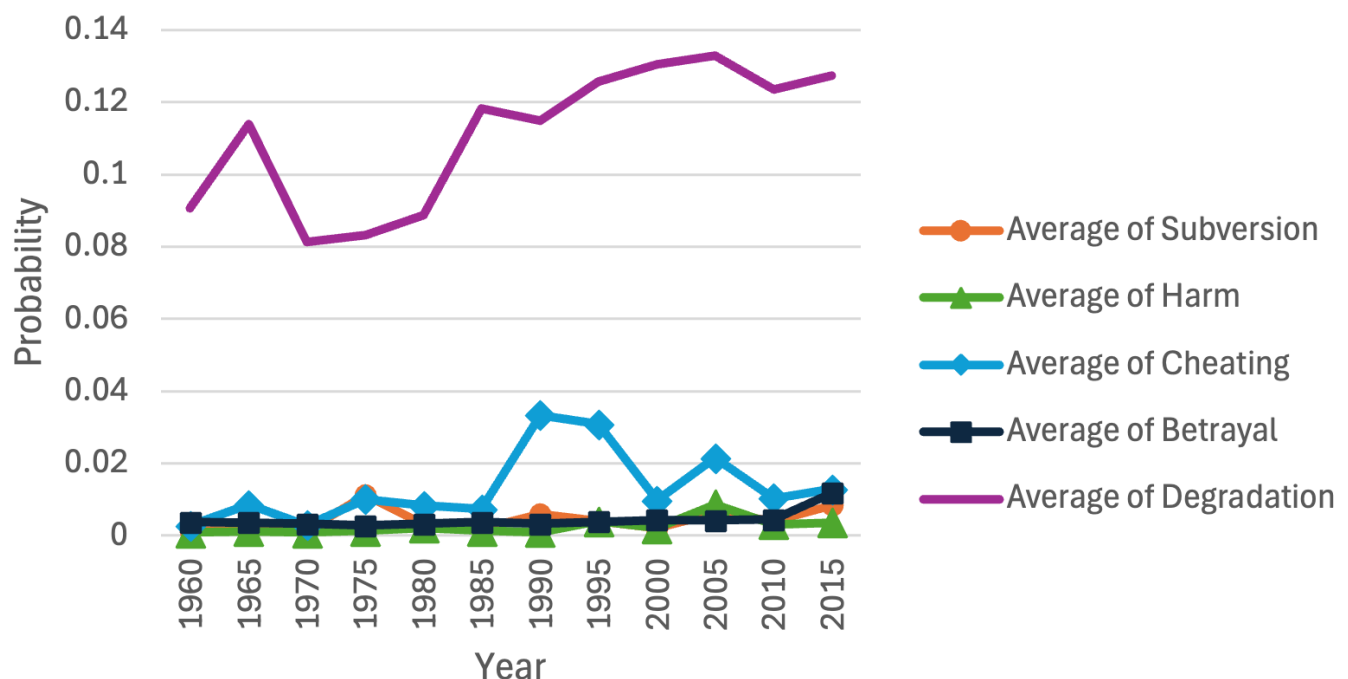


Figure 1. Average moral foundation probabilities of “Vices” across articles published in seven art journals from 1960–2015. Moral foundation probabilities of 5431 full texts derived from seven art journals were assessed by Moral Bidirectional Encoder Representations from Transformers (MoralBERT). We sampled from selected art journals every five years between 1960 and 2015, and the foundation probabilities of all articles in publication a given year were averaged; each year on the axis contains data for a single year, not the 5-year period between that year and the next year on the axis. Shown in the figure are the probabilities of five “Vices”: “Subversion,” “Harm,” “Cheating,” “Betrayal,” and “Degradation” (as per the Moral Foundations Theory). MoralBERT “Vice” probability scores range from 0 to 1. Each value on the line is representative of the average foundation probability for the corresponding “Vice” across all the art journal articles analyzed. “Degradation” was consistently the highest and exhibited the greatest magnitude of change. Orange, average probability of “Subversion”; green, average probability of “Harm”; light blue, average probability of “Cheating”; dark blue, average probability of “Betrayal”; purple, average probability of “Degradation.”

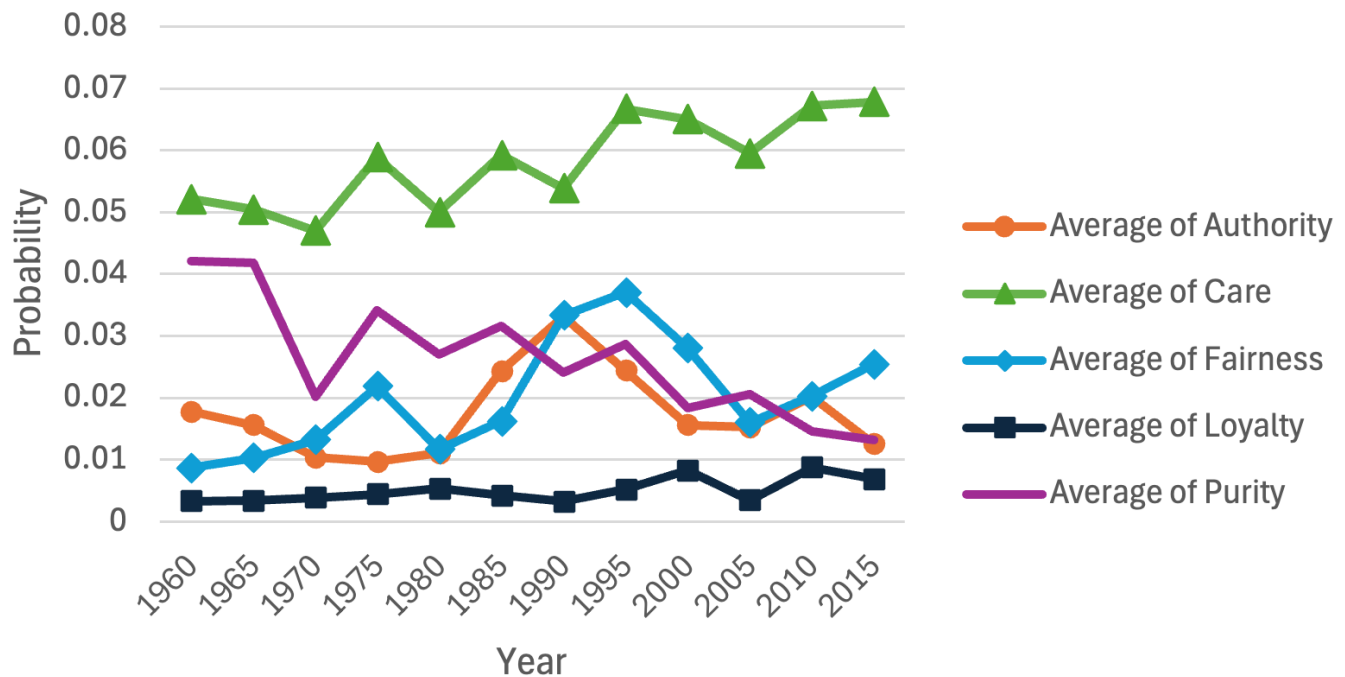


Figure 2. Average moral foundation probabilities of “Virtues” across articles published in seven art journals from 1960–2015. Moral foundation probabilities of 5431 full texts derived from seven art journals were assessed by MoralBERT. We sampled from selected art journals every five years between 1960 and 2015, and the foundation probabilities of all articles in publication a given year were averaged. Shown in the figure are the probabilities of five “Virtues”: “Authority,” “Care,” “Fairness,” “Loyalty,” and “Purity” (as per the Moral Foundations Theory). MoralBERT “Virtue” probability scores range from 0 to 1. Each value on the line is representative of the average foundation probability for the corresponding “Virtue” across all the art journal articles analyzed. “Care” exhibited an upward trend, while “Purity” decreased. Orange, average probability of “Authority”; green, average probability of “Care”; light blue, average probability of “Fairness”; dark blue, average probability of “Loyalty”; purple, average probability of “Purity.”

“Virtue”/“Vice” pairs, we found that “Care” was significantly greater than “Harm” ($t = 46.741$; $p < 0.0001$; $df = 5430$), and “Fairness” was significantly greater than “Cheating” ($t = 5.1628$; $p < 0.0001$; $df = 5430$). “Loyalty” and “Betrayal” were not significantly different ($t = 1.2631$; $p = 0.2066$; $df = 5430$). “Authority” was significantly greater than “Subversion” ($t = 9.5723$; $p < 0.0001$; $df = 5430$), while “Purity” was significantly less than “Degradation” ($t = -36.758$; $p < 0.0001$; $df = 5430$). The texts featured three of the five “Virtues” more frequently than their corresponding “Vices.” These results suggest that, when art critics write about art, their choice of language emphasizes the positive characteristics of the work over the negative.

DISCUSSION

Moral language in the text was limited. The average “Vice” probabilities for each year remained below 14%, and “Virtue” probabilities were less than 7%. “Care” and “Degradation” appeared more often in the texts compared to the other foundations. Both foundations increased over time, with “Care” exhibiting a 1.57% increase between 1960 and 2015, and “Degradation” rising by 3.67% across the period. Meanwhile, the “Purity” foundation decreased, falling from 4.21% to 1.31%.

We had expected all “Vice” and “Virtue” foundations to appear in a similar trend to one another, with all “Vices” exhibiting a net positive increase between 1960 and 2015, and all “Virtues” demonstrating a net decrease between

1960 and 2015. However, we had not expected that individual “Vices” and “Virtues” would display different patterns from one another, with some foundations contradicting the predicted direction of motion and others fluctuating or remaining stagnant. We found partial support for our hypothesis, most notably through the increase in “Degradation” and decrease in “Purity” across time.

The increase in the probability of “Degradation” corresponded with the negative trend in the sentiment of art criticism articles’ full texts (3). Meanwhile, the trend in “Care” mirrored that of the sentiment in art review titles, as both became more positive over time (3). This finding contradicted our hypothesis, which had assumed a decline in “Care” (a “Virtue” foundation) in alignment with the negative trend of the articles’ full texts described by Bao and Gao (3).

When we created linear trendlines for the foundations, the positive slopes of “Harm,” “Betrayal,” and “Degradation” supported the hypothesis, as we expected that the use of “Vices” would increase over time. In our hypothesis, we also proposed that the probabilities of “Virtues” would decrease. The negative slope of “Purity” provided evidence in favor of the hypothesis, while the positive slopes of “Care” and “Loyalty” contradicted the hypothesis.

The decline in the “Purity” foundation indicates that art critics gradually shifted away from writing about temperance, piety, and cleanliness in response to art. Additionally, art itself could have become more radical, as artists may have been less conscious about conforming to standards of “Purity.” This

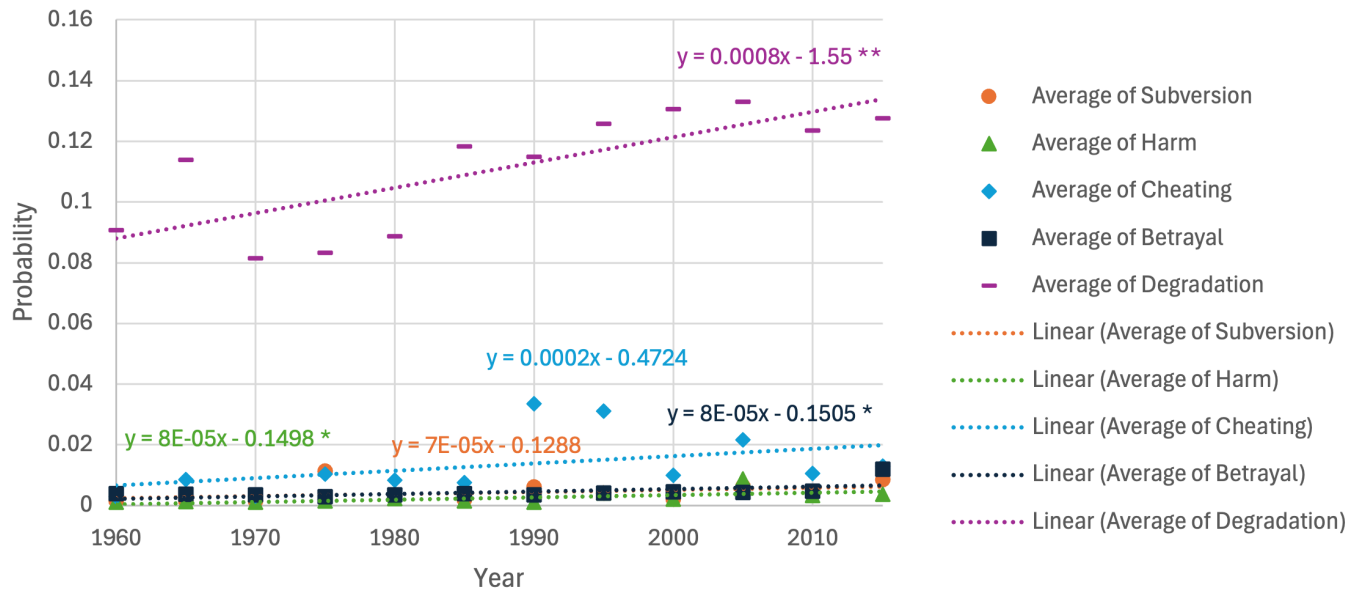


Figure 3. Scatterplot and linear trendlines for moral foundation probabilities of “Vices” across articles published in seven art journals from 1960–2015. Moral foundation probabilities of 5431 full texts derived from seven art journals were assessed by MoralBERT. We sampled from selected art journals every five years between 1960 and 2015, and the foundation probabilities of all articles in publication a given year were averaged. Shown in the figure are the probabilities of five “Vices”: “Subversion,” “Harm,” “Cheating,” “Betrayal,” and “Degradation” (as per the Moral Foundations Theory). MoralBERT “Vice” probability scores range from 0 to 1. Using Microsoft Excel’s Analysis Toolpak, we applied linear regressions to each of the “Vices.” “Harm,” “Betrayal,” and “Degradation” were positively correlated with time. The correlations between time and “Subversion” and “Cheating” were not significant. * $p < 0.05$, ** $p < 0.01$. Orange, average probability of “Subversion”; green, average probability of “Harm”; light blue, average probability of “Cheating”; dark blue, average probability of “Betrayal”; purple, average probability of “Degradation.”

finding is similar to results from a previous study (13). Gao and Jiang studied word frequencies in the same art journals as used in this study and across the same time period (13). From 1960 to 2015, the mention of ancient Greek and Roman art in the journals decreased (13). Through sculptures and images, the Greeks and Romans portrayed their gods (14). However, cultural preferences toward pragmatism led to a decrease in religious subject matter in art while simultaneously contributing to economic growth (15). The declining emphasis on religion disincentivized art critics from writing about religious art. Religious narratives are rich in the “Purity” foundation, as they tend to emphasize the values of temperance and chastity that characterize “Purity” (6). Thus, the inclusion of the “Purity” foundation in journal articles likely decreased when art critics’ attention shifted away from religious pieces.

The increase in “Degradation” across time could have resulted from the discussion of art that actively violated concepts of “Purity.” These art pieces contradicted the “Purity” foundation, which outlines the desire to be less carnal and more noble (6). During our time frame of analysis, articles such as “The State of Art History: Contemporary Art” shifted from discussing religious works to highlighting works less concerned with cleanliness and nobility, including Josephine Meckseper’s art piece, *The Complete History of Postcontemporary Art* (2005), which was constructed from city trash (16). The article aligns with the “Degradation” foundation, as it suggests a lack of cleanliness.

The “Degradation” foundation had higher probabilities in comparison to the other “Vices.” This pattern relates to the finding that contemporary art critics only address religion

when artists produce sacrilegious works of art, as art critics find such artwork more interesting (17). These scandals would align with the “Degradation” foundation. Outside of works of art that address religion in a satirical, insincere manner, art that functions in religious settings rarely appears in the academic discourse (17). We speculate that art critics dedicate more attention to sacrilegious artwork than to pieces that address the other “Vices.” Because of this difference in interest, “Degradation” exhibited the highest probabilities of all the “Vices.” Furthermore, “Degradation” displayed the highest magnitude of change relative to “Subversion,” “Harm,” “Cheating,” and “Betrayal.” This trend indicates that art critics gradually dedicated more focus to “Degradation” than to other “Vices.”

According to our linear regression, the “Care” foundation increased across time. This finding contradicts our prediction that moral “Virtues” would decrease. Nevertheless, the increase in “Care” can be explained with the role of curators in cultivating notions of “Care” and rendering exhibition spaces hospitable, particularly in response to times of crisis (18). For example, as part of the major art exhibition *documenta fifteen*, contemporary art collective Ruangrupa criticized systems of patriarchy, capitalism, and colonialism for promoting selfish individual ownership of items (18). Ruangrupa encouraged visitors to support one another by purchasing “Soli-Tickets” for those who could not afford to attend the exhibition (18). Additionally, Ruangrupa invited other art collectives to engage in the curatorial process, highlighting the power of collectivism over individualism (18). Together, the artists cared for one another and stood in solidarity against oppressive practices (18). Similarly, feminist philosophy in art has gained

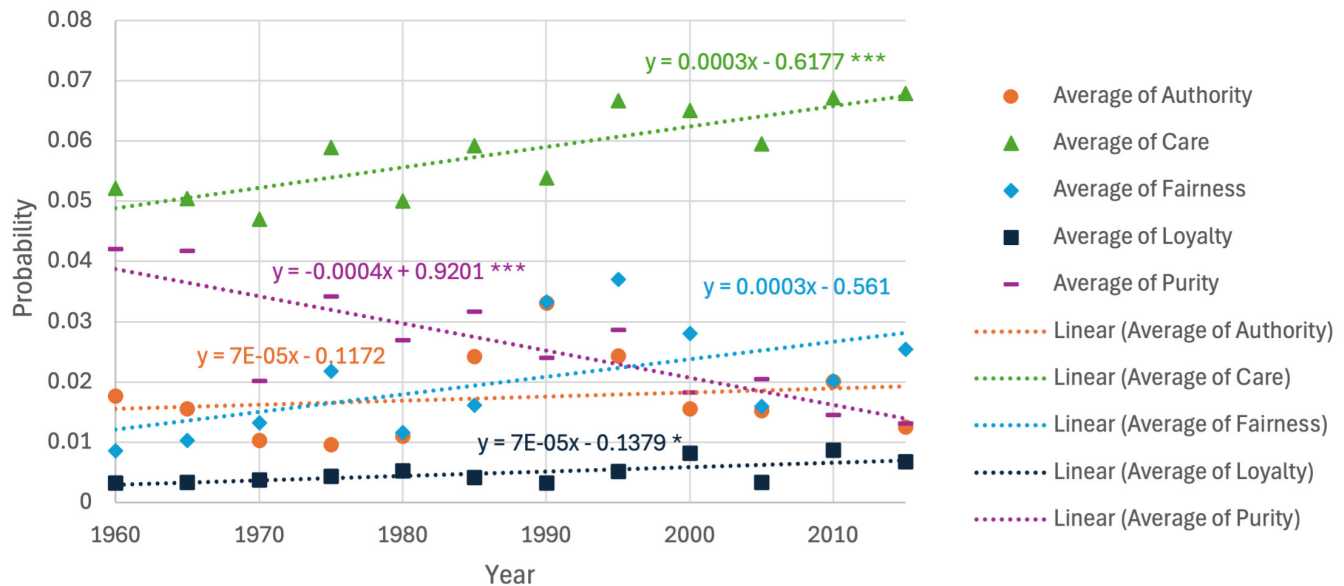


Figure 4. Scatterplot and linear trendlines for moral foundation probabilities of “Virtues” across articles published in seven art journals from 1960–2015. Moral foundation probabilities of 5431 full texts derived from seven art journals were assessed by MoralBERT. We sampled from selected art journals every five years between 1960 and 2015, and the foundation probabilities of all articles in publication a given year were averaged. Shown in the figure are the probabilities of five “Virtues”: “Authority,” “Care,” “Fairness,” “Loyalty,” and “Purity” (as per the Moral Foundations Theory). MoralBERT “Virtue” probability scores range from 0 to 1. Using Microsoft Excel’s Analysis Toolpak, we applied linear regressions to each of the “Virtues.” “Care” and “Loyalty” were positively correlated with time. Meanwhile, “Purity” decreased across time. The correlations between time and “Authority” and “Fairness” were not significant. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Orange, average probability of “Authority”; green, average probability of “Care”; light blue, average probability of “Fairness”; dark blue, average probability of “Loyalty”; purple, average probability of “Purity”.

prominence over time (19). Feminist philosophy draws from the “Care” foundation, emphasizing inter-connectedness and mutual aid (19). This approach promotes positive change for entire communities, rather than sole individuals (19).

MoralBERT was trained on language sourced from social media platforms, while most of the articles in our dataset were published before the emergence of social media (12). Given that the use of language changes over time, MoralBERT may not have accurately captured pre-social media language. Additionally, the casual tone of social media posts contrasts with the academic writing in art journals; MoralBERT may not have the capacity to make predictions on formal language. Therefore, finetuning MoralBERT for art journals in the time period of interest would likely improve the accuracy of results.

Future studies should incorporate the journal *Frieze* into the dataset, as *Frieze* was present in the sentiment analysis study (3). This additional data may impact the average moral foundation scores for each year and consequently affect the trends in moral language across time. It would also allow for more accurate comparison with the sentiment analysis trends due to the adoption of the same datasets.

The art journals did not organize works of art, artists, and art critics by their country of origin. As a result, we could not analyze correlations to country-specific events. Our dataset only included English articles, as MoralBERT cannot perform analysis in other languages. Future studies could include art journals that are not written in English and have been published in other countries, thereby encompassing a wider demographic; this task would require training MoralBERT in non-English languages. Additionally, the original MFT does not represent an exhaustive list of all facets of human morality (6). Recently discovered foundations, such as “Equality,”

“Proportionality,” and “Liberty,” could offer more nuance to our analysis of art criticism articles (20).

In conclusion, using MoralBERT, we evaluated trends in art journal articles’ moral language from 1960 to 2015. Changes across time were minimal for both “Vices” and “Virtues.” As indicated by an analysis of the most frequent score ranges, all foundation probabilities tended to be low. Art critics may not use much moral language in their writing. Alternatively, since MoralBERT was not fine-tuned for art journal articles, it may not have been able to adequately detect moral language in this context. Nevertheless, we found that while the probability of “Purity” appearing in the texts decreased across time, the “Degradation” and “Care” foundation probabilities increased across time. Most “Virtues” appeared more frequently than their corresponding “Vices.” Therefore, art criticism content from 1960 to 2015 exhibited more positive moral characteristics than negative overall. These results offer insight into art critics’ use of moral language, along with its evolution in the context of wider art movements.

MATERIALS AND METHODS

Data collection

Our data consisted of 5431 full-text articles from seven art journals, namely *Afterall: A Journal of Art, Context and Enquiry*; *Art Journal*; *Grey Room*; *Leonardo*; *October*; *The Art Bulletin*; and *The Journal of Aesthetics and Art Criticism*. These same journals contained trends in sentiment (3). We were unable to obtain articles from *Frieze*, the eighth art journal from the sentiment analysis study, due to their paid membership policy (3). Articles were sampled every five years from 1960 to 2015 in accordance with Bao and Gao’s methodology; articles published between the five-year

Vices	Harm	Cheating	Betrayal	Subversion	Degradation
	0.00–0.01	0.00–0.01	0.00–0.01	0.00–0.01	0.02–0.03
Virtues	Care	Fairness	Loyalty	Authority	Purity
	0.02–0.03	0.00–0.01	0.00–0.01	0.00–0.01	0.00–0.01

Table 1. Most frequent probability score ranges for each moral foundation. Moral foundation probabilities of 5431 full texts derived from seven art journals were assessed by MoralBERT. We sampled from the journals every five years between 1960 and 2015, and the foundation probabilities of all articles in publication a given year were averaged. The probabilities ranged from 0 to 1, and we divided all the scores into bins of 0.01. The table contains the most frequent range of the probabilities for each foundation. In the ranges shown, the lower bound is exclusive and the upper bound is inclusive. The most frequent score ranges for “Care” and “Degradation” were between 0.02 and 0.03. The rest of the foundations stayed in the 0.00–0.01 range.

endpoints were not included (3). Of the seven art journals, *Art Journal*, *The Art Bulletin*, and *The Journal of Aesthetics and Art Criticism* had publications in circulation by 1960. Meanwhile, *Afterall*, *Grey Room*, *Leonardo*, and *October* did not publish their first issues until after this year. When we collected articles from each of these four journals, we began with the first publication year that was a product of five. Thus, our dataset contained articles beginning in the year 2000 for *Afterall*, 2000 for *Grey Room*, 1970 for *Leonardo*, and 1980 for *October*. After a journal’s first publishing date, there were no other missing years. In total, 130 articles were gathered from *Afterall*, 961 from *Art Journal*, 111 from *Grey Room*, 1679 from *Leonardo*, 312 from *October*, 963 from *The Art Bulletin*, and 1275 from *The Journal of Aesthetics and Art Criticism*. The full text was extracted via Constellate, a platform for text analysis (21).

Data analysis

We evaluated moral values in the texts with MoralBERT (22). MoralBERT is a language model that relies on the BERT-base-uncased pretrained sequence classifier to analyze moral sentiment in social media posts (12). In pre-existing datasets, human annotators had labeled social media posts from Twitter, Reddit, and Facebook as reflecting particular moral foundations or as possessing non-moral content (12). MoralBERT was trained on the text of the social media posts, along with the human labels (12). As part of its in-domain classification tasks, the model was trained and tested on all three corpora (12). For out-of-domain classification, the model learned from two of the corpora and made new predictions on the third corpus (12). The authors created two versions of MoralBERT—one with aggregated training and one with domain-adversarial training (MoralBERT_{adv}) (12). They evaluated MoralBERT’s performance with F1 binary and F1 macro scores. F1 binary scores reflected the model’s accuracy in predicting true positives for moral content (12). F1 macro scores indicated the model’s accuracy in predicting true positives and true negatives across moral, non-moral, and neutral content (12). For in-domain predictions, the two models achieved average F1 scores between 11% and 32% higher than those of other approaches (12). These methods included models reliant on lists of words, models that converted text to vectors, and the use of Generative Pre-trained Transformer 4 (GPT-4) to assign moral labels to social media posts without first training on human annotations (12). In out-of-domain predictions, MoralBERT_{adv} achieved similar average F1 binary and macro scores as GPT-4, with 1%

and 2% differences, respectively (12). Thus, MoralBERT_{adv} is comparable to current large language models for out-of-domain predictions. For this reason, we were interested in applying MoralBERT_{adv} to the new domain of art critique journals.

MoralBERT produced predicted probabilities for the moral foundations (12). The model outputted values between 0 and 1, with a higher value indicating that a greater portion of the text expressed a moral foundation. Additionally, MoralBERT was derived from the original MFT and Moral Foundations Questionnaire (MFQ-1) (12). The authors of the original theory clarified that, prior to 2012, they referred to “Sanctity/Degradation” as “Purity/Sanctity” with a focus on “Virtue” principles (6). MoralBERT produced a “Purity” score, which corresponded to the “Sanctity” foundation. Overall, the “Virtues” for MoralBERT were “Care,” “Fairness,” “Loyalty,” “Authority,” and “Purity” (12). The “Vices” were “Harm,” “Cheating,” “Betrayal,” “Subversion,” and “Degradation” (12). The maximum token length was assigned a value of 512 for all articles, as the BERT model could process no more than this number of tokens (11). During the preprocessing stage, all texts were lowercased. The WordPiece algorithm performed tokenization by comparing each word to a pre-learned vocabulary list (23). Reading from the beginning of the word, the algorithm found the longest corresponding subword in the vocabulary and split the word (23). For the remainder of the original word, the algorithm once again located a subword, continuing until the entire word was tokenized (23). If the algorithm could not locate a subword, the entire word was labelled as unknown (23). In cases where the algorithm located a subword, but the rest of the word contained no more subwords, the entire word was once again classified as unknown (23).

Statistical tests

For each foundation, scores across all articles and years were divided into bins of 0.01 from 0 to 1. The frequency of a given bin was the number of scores contained in the range of that bin, divided by the total number of scores. The bin containing the highest frequency of each moral foundation was calculated. In another test, Microsoft Excel’s Analysis Toolpak was used to produce linear regressions between the average moral foundation values and time. Finally, paired two-sample t-tests were performed between each of the “Virtue”/“Vice” pairs across the period from 1960 to 2015.

Received: March 3, 2025
Accepted: August 8, 2026
Published: October 23, 2026

REFERENCES

1. "Art Journal." *JSTOR*. <https://www.jstor.org/journal/artj>. Accessed 28 Jan. 2025.
2. Klöckner, Christian Andreas and Laura K. Sommer. "Visual art inspired by climate change—An analysis of audience reactions to 37 artworks presented during 21st UN climate summit in Paris." *PLOS ONE*, vol. 16, no. 2, 19 Feb. 2021, e0247331. <https://doi.org/10.1371/journal.pone.0247331>.
3. Bao, TianCheng and Junhui Gao. "Research on American Contemporary Art Review Based on Sentiment Analysis Technology." *Proceedings of the 2019 International Symposium on Signal Processing Systems*, 20 Sept. 2019, pp. 136–41. <https://doi.org/10.1145/3364908.3365293>.
4. Feldman, Ronen. "Techniques and applications for sentiment analysis." *Communications of the ACM*, vol. 56, no. 4, 1 Apr. 2013, pp. 82–89. <https://doi.org/10.1145/2436256.2436274>.
5. Nemes, László and Attila Kiss. "Social media sentiment analysis based on COVID-19." *Journal of Information and Telecommunication*, vol. 5, no. 1, 14 July 2020, pp. 1–15. <https://doi.org/10.1080/24751839.2020.1790793>.
6. Graham, Jesse, et al. "Chapter Two - Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism." *Advances in Experimental Social Psychology*, vol. 47, 8 Apr. 2013, pp. 55–130. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>.
7. Devereaux, Mary. "Moral Judgments and Works of Art: The Case of Narrative Literature." *The Journal of Aesthetics and Art Criticism*, vol. 62, no. 1, 2004, pp. 3–11. <https://doi.org/10.1111/j.1540-594X.2004.00130.x>.
8. Graham, Jesse, et al. "Liberals and conservatives rely on different sets of moral foundations." *Journal of Personality and Social Psychology*, vol. 96, no. 5, 2009, pp. 1029–1046. <https://doi.org/10.1037/a0015141>.
9. Clifford, Scott and Jennifer Jerit. "How Words Do the Work of Politics: Moral Foundations Theory and the Debate over Stem Cell Research." *The Journal of Politics*, vol. 75, no. 3, July 2013, pp. 659–71. <https://doi.org/10.1017/S0022381613000492>.
10. Hopp, Frederic R., et al. "The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text." *Behavior Research Methods*, vol. 53, 14 July 2020, pp. 232–246. <https://doi.org/10.3758/s13428-020-01433-0>.
11. Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186. <https://doi.org/10.18653/v1/N19-1423>.
12. Preniqi, Vjosa, et al. "MoralBERT: A Fine-Tuned Language Model for Capturing Moral Values in Social Discussions." *Proceedings of the 2024 International Conference on Information Technology for Social Good*, 4 Sept. 2024, pp. 433–442. <https://doi.org/10.1145/3677525.3678694>.
13. Jiang, Yanhui and Junhui Gao. "Word Frequency Analysis of American Contemporary Art Reviews from 1960 to 2015." *Open Journal of Modern Linguistics*, vol. 8, no. 4, Aug. 2018, pp. 87–98. <https://doi.org/10.4236/ojml.2018.84010>.
14. Osborne, Robin and Caroline Vout. "Art and Religion in Ancient Greece and Rome." *Oxford Research Encyclopedia of Religion*, 9 May 2016. <https://doi.org/10.1093/acrefore/9780199340378.013.81>.
15. Kavolis, Vytautas. "Art Content and Economic Reality." *The American Journal of Economics and Sociology*, vol. 24, no. 3, Jul. 1965, pp. 321–328. www.jstor.org/stable/3485124.
16. Smith, Terry. "The State of Art History: Contemporary Art." *The Art Bulletin*, vol. 92, no. 4, 1 Apr. 2014, pp. 366–383. <https://doi.org/10.1080/00043079.2010.10786119>.
17. Elkins, James. "From Bird-Goddesses to Jesus 2000: A Very, Very Brief History of Religion and Art." *Thresholds*, no. 25, 2002, pp. 76–83. https://doi.org/10.1162/thld_a_00382.
18. Ren, Julie, et al. "Crisis and Collectives Shaping Art Events: From Events that Matter to Events that Care." *Third Text*, vol. 37, no. 5–6, 12 Jan. 2024, pp. 601–615. <https://doi.org/10.1080/09528822.2023.2297626>.
19. Millner, Jacqueline. "Care ethics and contemporary art: Imagining and practising care." *Thesis Eleven*, vol. 183, no. 1, 23 Oct. 2024, pp. 103–118. <https://doi.org/10.1177/07255136241284659>.
20. Dobolyi, David. "Home." *Moral Foundations Theory, YourMorals*, Sept. 2024. <https://moralfoundations.org>. Accessed 26 Dec. 2025.
21. Chang, Emerson. *Seven Art Journals*. Constellate, 30 Sept. 2024. constellate.org/dataset/27f3eefa-1d21-f3c4-7405-fee424b729e7. Accessed 30 Sept. 2024.
22. vjosapreniqi. "Predict_mft_scores_from_the_MoralBERT_weights.ipynb." *MoralBERT*, commit 176ade0, 2024. github.com/vjosapreniqi/MoralBERT/blob/main/MoralBert/Predict_mft_scores_from_the_MoralBERT_weights.ipynb.
23. "WordPiece tokenization." *Hugging Face*, huggingface.co/learn/llm-course/en/chapter6/6. Accessed 10 Aug. 2025.

Copyright: © 2026 Chang and Mahmoud. All JEI articles are distributed under the Creative Commons Attribution Noncommercial No Derivatives 4.0 International License. This means that you are free to share, copy, redistribute, remix, transform, or build upon the material for any purpose, provided that you credit the original author and source, include a link to the license, indicate any changes that were made, and make no representation that JEI or the original author(s) endorse you or your use of the work. The full details of the license are available at <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>.