Measuring Abstraction Levels of Sculptural Objects

Gordon E. Lyon[1] and Merritt R. Lyon[2]

[1] Independent artist, 390 Russell Ave., Gaithersburg, MD 20877 <gelyon@verizon.net>

[2] Department of Statistics, Columbian College of Arts and Sciences, George Washington University, Rome Hall 801 22nd St. NW, 7th Floor, Washington, DC 20052

NOTE to READERS: This is the manuscript as accepted. Several minor adjustments from referees were incorporated directly into the text by the publisher, De Gruyter Brill. We did add some font adjustments to the typeset version for notational consistency. Lack of these changes in this version should cause no problems. See:

Lyon, G. E., & Lyon, M. R. (2024). Measuring Abstraction Levels of Sculptural Objects. ***Art & Perception*** (published online ahead of print 2024). https://doi.org/10.1163/22134913-bja10059

Abstract

Freestanding sculpture can be scored for abstraction using order scales such as the 3-rank 'realistic,' 'mixed,' and 'abstract.' Subjective but invaluable, the ranks support ordering but not arithmetic. Numeric limitations can be addressed via information-theoretic metrics that measure collective uncertainty about a sculpture; to this end 60 participants view 25 images of sculpture via an internet questionnaire. They (i) score each depicted object and (ii) type a caption stating what impression the object evokes. Captions for the image are simplified to their English simple subject, sorted into classification categories and counted. A metric converts an image's categories and counts into viewers' uncertainty. The investigation examines 4-rank and 5-rank scales plus metrics CC (category count), H (entropy) and EC ≡ 2^H (equalized category count). Medians of scores or uncertainties reduce variability as appropriate.

Comparisons of scores versus uncertainties show the two generally correlate well. Results also highlight inherent design tradeoffs. Scale/metric pairings affect both ranks (which should be statistically distinct) and rank/uncertainty correlations. Two good combinations are [4-rank; median CC] and [5-rank; median EC]. Metric CC is easy to compute and correlates slightly better than EC, although the latter supports five evenly separated ranks. Shifting focus to image scatterplots, the pair [4-rank median score; CC] yields a very strong correlation. Once uncertainties are linked to ranks to "calibrate" them, one gets a quantitative sense of the rank order scale. The augmented framework offers the speed and ease of scoring along with valid numeric estimates of uncertainty unavailable from ranks alone.

Keywords: abstraction level; measurement; metrics; quantitative; sculpture; subjective

1. Introduction

All art involves some degree of abstraction (Zimmer, 2003), but most artworks are not completely abstract. How abstract are they? Viewers reliably make group estimates of abstraction levels. This investigation examines these estimates through the lens of information theory and its metrics, the idea being to correlate subjective group estimates with more objective statistical measures. To do this it is necessary to restrict the scope of study to something manageable. The chosen art subfield comprises *sculptural objects,* freestanding 3-dimensional pieces such as A. Rodin's 'The Thinker' or B. Hepworth's 'Pelagos'.

Abstraction and Sculpture
Abstract or non-representational sculpture arose early in the twentieth century. Rodin had freed sculpture from architecture, but Brancusi and admirers of Cezanne would take it into realms of the non-representational (Read, 1964; Curtis, 1999). Advance was rapid. Brancusi sculpted his semi-abstract 'The Kiss' in 1907 and a more abstract 'Two Penguins' in 1911-14. By 1928 his 'Bird in Space' series (1923-) had famously clashed with the US Customs Service (Brancusi, 1928). In 1914, Picasso constructed a cubist 'Still Life' (Tucker, 1974). V. Tatlin, inspired by a 1913 visit to Picasso, initiated what became Russian Constructivism (~1915-1930), a major art movement that spread sculptural abstraction internationally (Lucie-Smith, 1987; Curtis, 1999).

Sculptural Objects
Sculptures in the first half of the twentieth century were mostly objects (see the illustrations in Read's (1964) book, *Modern Sculpture*). The 1960s saw sculptors increasingly emphasizing audience encounters: installations; performances; light or sound shows (Causey, 1999). Sculptor/theorist Robert Morris wrote in 1966 that "...The object is but one of the terms in the newer aesthetic" (Moszynska, 2013, p. 9). Clement Greenberg objected, but Morris's remark prevailed (Idem., p. 12). A widely-read essay by Rosalind Krauss (1979) proposed what, beyond objects, post-modern sculpture might entail. Yet forty-five years later, sculptural objects remain popular and enjoy an expanded presence. For example, contemporary artist Alice Aycock (1946-) may appropriate floor, walls and ceiling for an installation. At other times she reifies a weather model as a huge, freestanding tornado-like object. Regarding a wider presence, N. Atkinson (2015) remarks in *Craft for a Modern World* that contemporary museum-level craft objects need no longer be tightly focused on immaculate craftsmanship. Many craftworks today evoke a rich variety of distinct associations characteristic of the abstract sculpture they are.

Public Response
Public appreciation of art—especially abstract works—is never guaranteed (Pelowski, et al.,

2017). A fraction of any population will likely not value abstract art (Wilson, A.S. and Cupchik, G. (1992)). Nonetheless, Gortais (2003) envisions back-and-forth interactions between an artistic formulation and the public, so "..Perhaps the work of art will enable the public to enjoy, in turn, a subjective artistic experience." Ideally, this dialog evolves positively over time. But not always: R. Hughes (1988) relates how sculptor R. Serra and the NYC public got radically out of synchronization over his immense outdoor sculpture, *Tilted Arc*. Viewers participating in this study are no different, even though their educational cross section mirrors prosperous areas of the US or Canada. Forty percent report "usually" or "always" *disliking* abstract sculpture. As discussed later, *dislikers* cause no problem in this investigation, but they constitute a factor worth considering in experiment design.

Measurements, Technology, Models

Data in this study come from an internet group completing a questionnaire. Participants evaluate images of sculptures. Empirical studies of aesthetics go back at least to G. Fechner (1801-1887), who stressed *experiments* and developed techniques for measuring aesthetic preference. Methods herein—rank selection (choice) and typed response (production)—echo Fechner. In the mid 1950s Daniel Berlyne began efforts that would reinvigorate experimental aesthetics. In roughly the same period, Gestalt psychologist Rudolf Arnheim published his influential *Art and Visual Perception* (Zimmer, 2003). Technology too was improving—rapidly. The advent of mini-computers, e.g., PDP-8 in 1965, enabled academic groups to conduct human performance experiments under flexible yet precise computer control. More recently, computers have provided fast internet service, medical imaging and crowdsourcing; all these expand possibilities for aesthetic experiments.

A viewer/artwork model often accompanies an aesthetic experiment. Takahashi (1995) distinguishes two approaches. The first focuses on the *art object*, seeking to identify visual factors that influence aesthetic judgments. For example, Cotter, et al. (2017) show viewers preferring curvature to angularity. The second emphasizes that "...what we know routinely influences what we see..." (Luypan, 2017). Verbal cues suggest cognition affects art perception, although distinct participant backgrounds yield different results (Ibid.). These points are valuable, but discussion here focuses primarily on *correlations between responses* (subjective and objective). Nonetheless, participants—as generators of responses—cannot be factored out. Informal perspectives of viewer-artwork interactions are useful later, when reflecting on what has been observed.

An Abstraction Framework

The continuum of abstraction, an ordered progression of ranks, occupies a pedagogical and conceptual role within fine art (Davidson 1985; Ocvirk, et al. 2006). Realistic art objects sit at one end with highly abstract, non-representational works at the other. The simple continuum framework adapts well. For example, Pihko, et al. (2011) place artworks into a hierarchical table whose five classification ranks range from I (most representational) to V (extremely abstract), whereas Uusitalo, et al. (2012) choose a nine-point (1—9) scale. In all cases, viewers score artworks subjectively (Jamieson, 2004).

Any improvement to the continuum would sharpen insights for inquiries such as, “What do viewers' perceived levels of abstraction signify?” and “Beyond an ordering, how do abstraction levels relate?” These questions have implications for design:

> “How can the theory and techniques of traditional visual arts help beautify modern technology outputs and products and enhance their usability?” (Zhang et al. 2012)

Knowing ambiguity levels along the continuum would help. The hypothesis here is that abstraction scores and viewers' collective uncertainty (ambiguity) about an artwork correlate. Since uncertainty is measurable statistically, this linkage could add quantitative attributes to the continuum.

But how are viewers' collective uncertainties captured? Wilson (2012) classifies and tallies participants' single-word responses to stimulus images (photomicrographs of stained cells). Each image becomes characterized by its partition P, a ensemble of word frequencies. Logarithms of P's largest frequencies determine printed sizes of corresponding words in the image's “word cloud,” a visual summary of viewers' impressions.

The approach here builds on Wilson's, although each participant responds *twice* to an artwork image: (1) a subjective abstraction score '1' to '5'; (2) a short typed caption that is first simplified, then classified and tallied in the image's partition, P. Word captions are central to the approach—their variety reflects uncertainty among participants about what a work evokes (Wilson 2012; Sidhu, et al. 2018; Uusitalo 2012; Lyssenko 2016). The artwork may itself be multi-representational, which further increases response variability (Tormey and Tormey, 1983). In lieu of Wilson's sized-word visual cloud for assessing an image, several information-theoretic metrics applied to *P* give distinct—but related—uncertainty measurements. Uncertainties and scores are then compared.

2. Materials and Methods

Images

Digital Library's JStor/ArtStor archive provide a starter set of 57 digital images of sculptures. (Note: ArtStor is now folded into JStor. See Appendix B for artwork titles and alternative sources. JStor links allow thumbnail views for everyone.) Image sizes range *circa* 750x550 to 1500x900 pixels, which display clearly on monitors and laptops. Depicted sculptures show no noticeably degraded workmanship. With one exception, the works date from 1860 to the present—modern and contemporary art periods. All sculptors have national or international standings.

Images are screened to meet a project requirement that each photograph clearly emphasize its art object (Hayn-Leichsenring, 2017). Professional digital photographs of sculptures from museums work well: Most are clear and carefully composed against neutral backgrounds (cf. Fig 1). To encourage seeing the works as single *objects*, image selection excludes multi-element sculptures such as Remington's “Coming through the Rye” or Hepworth's “Four Figures Waiting.” Images of art installations are also barred.

Participants

Sixty US adults participated interactively via Amazon's service MTurk (Mechanical Turk). All participants held MTurk's Master qualification for reliable performance: Loepp and Kelly (2020) recommend this to promote sample quality. Additionally, participants had to have successfully completed at least 95% of their previous MTurk tasks. MTurk on-line documentation for working with Google Forms (the questionnaire vehicle) suggested collecting participants' MTurk Worker ID (anon., 2017). This discouraged retaking the questionnaire. There were no ID repetitions. Participants were asked no other control information (Agley, 2022).

Self-identified genders were 30 *male*, 29 *female* and one *other*. Ages ranged from 24 to 68 years old. The age median was 38.5 years, the mean, 41.8. Thirty-five percent of the group (21 respondents), held a 4-year college degree or equivalent. Thirty percent (18) declared little or no formal training beyond high school. Another 25% (15) held 2-yr. associate degrees, multi-year apprenticeship certificates, etc. Ten percent (6) had graduate or professional degrees. Asked how many art exhibits (in libraries, fairs, galleries, museums, artists' studios, etc.) they visited in an average year before Covid-19, 75% said one or two, 20% none and 5%, 3-8. At the questionnaire exit participants were asked: “Having viewed 25 images, would you say you like abstraction in sculpture?” They responded: 43.3% (26)--often, but not always; 33.3% (20)--sometimes, but often not; 16.7% (10)--yes; 6.7% (4)--no. All participants agreed that their (anonymous) responses could be shared with others.

Experimental Procedure

The set of images should represent ranks of the continuum as uniformly as possible. For this, one pretests and culls a deliberately oversized starting set. Five local participants who did not take the main questionnaire each scored the 57 initial images for abstraction using ordered ranks of '1' (realistic) to '5' (abstract). The median score for each image placed it in one of five ranks. From the ten to 14 images in each rank, five were then randomly selected to build a final set of 25. Because scores were later redetermined in the main questionnaire, one had to settle for likely having improved coverage. Final image rankings appear in Appendix B. It is important that the number of test images (25) be large enough to establish rank correlation confidences for *image* median scores vs. uncertainties. A later section, *Artwork-Centered Statistics, discusses* this.

To help participants adjust to scoring images with the '1'-'5' scale, the questionnaire provided three example image rankings from the culled images. Scores were defined as:

1-realistic (a detailed, consistent representation of the world)
2-tending realistic (less reliable detail than #1)
3-mixed (the topic may be hidden somewhat)
4-tending abstract (non-representational elements dominate)
5-abstract (truly non-representational; nothing familiar)

As further help, an abbreviated scale repeated beneath every test image. No information on artist, title or year of creation was shown (Franklin, et al. 1995). Five extra internet participants, paid $5, established timing estimates and confirmed workability of the final design. Self-paced, the questionnaire was made available across the USA during a two-hour weekday period starting at 10:30am EST. Average completion time was under 23 minutes and participants earned $6. There were no collection problems and all responses were useable. The questionnaire presented each participant 25 images in random order. Sixty participants (a) scored each depicted work '1' to '5' and (b) typed a caption of 36 or fewer characters describing whatever impression the image evoked for them. They did not have to identify anything, although many did. This yielded 1500 (score, caption) pairs for analysis.

Metrics, Classifications and Analysis
CC is a simple, robust metric that ignores partition frequencies. It simply counts categories, i.e., different caption types . This lends it an ease of application, e.g., *CC* works even when limited participation render frequency data incomplete or suspect. *CC's* quick and intuitive nature renders it well-suited for demonstrations. Expect *CC* to be higher than a weighted category count (EC, below).

Metric ensemble entropy H (Shannon, 1949) assumes P has representative frequencies from adequate sampling. Because H uses partition information fully, it enjoys another form of robustness. Any misclassified caption is discounted in $H(P)$ calculations because the caption will have a low frequency weight. However, H is not intuitive within an art community more attuned to visions of ruin associated with physical entropy. Observe that Shannon's H is the root formulation from which physical entropy derives (E.T. Jaynes, 1957).

The third metric, $EC \equiv 2^{H(P)}$, melds *CC's* intuitive counting with *H's* weighted categories. Imagine a new partition, P^*, with $2^{H(P)}$ categories such that $H(P^*) = H(P)$; this holds only when all category frequencies of P^* are equal. *EC* thus converts P into P^*, with its categories 'equalized' for counting. Comparisons using *EC* are as straightforward as with *CC*: It is one partition's *EC* versus anothers. Ideally, *EC* 's frequencies offer more balanced comparisons (infrequent captions can distort *CC*). Note that $EC(P) \equiv 2^{H(P)} \leq CC(P)$, since P^* is the partition of fewest categories having entropy $H(P)$.

The MTurk + Google Forms combination returns questionnaire results via one large table. Editor and spreadsheet operations suffice for classifying captions. Captions are reduced to English grammar *simple subjects* via inspection and text editing. Expletives, exclamations and the like are taken literally, as long character strings (see Appendix A). Reduced captions for an image are then sorted into partition categories where they are counted. An *optional* step taken prior to removing labels merges categories via tailored label-equivalence rules such as *cat ≡ feline*. Overall, the classification scheme is designed to be (i) simple, (ii) transparent, (iii) modular and (iv) reproducible.

Each image *i* is described by 60 scores and a partition of frequencies, P_i. Together they form a set, $\{(score, P_i)\}$, of 60 intermediate data pairs. Given 25 images, there are 60 x 25 = 1500 intermediate pairs. A metric *M* (= *CC, H or EC)* is then applied to all intermediate pairs, converting them to a dataset of 1500 x-y points, $\{(score, M(P))\}$. *M* can be one of three different operators, so there are three datasets.

3. Results

Image scores are first checked as to whether, for each image, participant-assigned scores are similar (cluster together) or scatter haphazardly over the '1' to '5' scale. Clustering indicates some overall participant consonance in determining scores. Assessed by an intraclass correlation test (ICC), viewer scoring performance is “moderate/good,” with consistency = *0.73* and agreement = *0.72.*

A Score-Focused Perspective.

The datasets display well in the cartesian format of Figs 2-5, which help screen for scoring/metric combinations. The several Box plots in each figure summarize image uncertainty measurements (y axis) for the five ordered ranks ('1-realistic' to '5-abstract' on the x-axis), the exception being composite rank '2-3' in Fig 3, discussed below. The plots show quartile confidence intervals. Widths for plotted boxes are proportional to the square root of the number of responses, which are 306 ('1-realistic'), 273 ('2-tending realistic'), 253 ('3-mixed'), 289 ('4-tending abstract') and 379 ('5-abstract'). Two inner quartile boxes delineate the central half and show the median. Each side notch indicates roughly a 95% confidence interval. Since notches in the figures do not overlap, all medians are likely distinct (McGill, 1978). Whiskers indicate data extremes.

/Figures 2 and 3 about here/

Fig 2 depicts Box plots for metric *M = CC.* A Spearman rank correlation for this dataset is rho = (.698)[1500], $p < 2.2e\text{-}16$, which has a rule-of-thumb interpretation of “moderate/strong.” Medians for *CC* show a strictly monotone relation with scores, although median confidence limits are just barely separated for ranks '2' and '3'. Merging '2' and '3' for a 4-rank scale improves separation in Fig 3 but costs an abstraction rank. Rho is essentially unchanged, dropping a tick to (.696)[1500], $p < 2.2e\text{-}16$.

Metric H in Fig 4 uses all frequency information. The scale is 5-rank. H median separations are good, but overall, rho declines mildly: (.658)[1500], $p < 2.2e\text{-}16$. The correlation is "moderate." $EC = 2^H$ (Fig 5) also separates medians better than *CC.* Because *EC* is a strictly monotone transformation of H, its Spearman rank correlation is also .658. (Not shown: For both H and EC, merging scores '2-3' for a 4-rank scale causes a tiny decrement; rho = (.656)[1500], $p < 2.2e\text{-}16$.) Figs 2-5 demonstrate a positive correlation between *raw scores* and *median uncertainties.*

/Figures 4 and 5 about here/

Median uncertainties quantify ranks of the continuum. Take for example Fig 5, which shows an invertible relationship between median *EC* (denoted $\underline{EC}$) and scores. $\underline{EC}$ in Fig 5 is 3.74 ('1-realistic'), 11.36 ('2-tending real'), 14.30 ('3-mixed'), 18.36 ('4-tending abst.') and 23.34 ('5-abstract'); units are *equalized categories.* The values encourage quantitative comparisons among continuum ranks, e.g.: uncertainty *EC* varies by a factor of 6.2 across the five ranks; moving from rank '4' to '2' cuts $\underline{EC}$ by 40%; ambiguity for rank '5' is considerable (23.3). These statements about ambiguity (uncertainty) provide tangible facts not available from rank descriptions. As always, such measurements reflect scoring, metric, classification and participants.

Artwork-Centered Statistics

Partitions and metrics certainly generate a miscellany of information about individual sculptures. Moloy-Nagy's abstract work *Nickel Construction* (1921) has $CC = 48$ $(EC = 43.8)$; it is very ambiguous (see Fig 1, #5). Opposite on the continuum, Rosa Bonheur's realistic *Lion* (1880) has a low $CC = 7$, with 50 (out of 60) captions classified in category *lion* (Fig 1, #1).

/Figure 6 about here/

A succinct overview of images is available from scatterplots of image median scores versus an uncertainty metric. Using median scores reduces variation and dataset pairs drop from 1500 to 25. The dataset is small but a Stuart-Kendall tau-c rank correlation works well (see Bonett and Wright (2000), Table 1, etc. for details). Results for H and EC will again be the same, so they are reported together:

	scoring (median)	
metric	4-rank	5-rank
CC	.811	.768
EC/H	.730	.720

Figs 6 displays the 25 test images using a (merged)` '2-3' rank that worked well earlier with metric CC (cf. Fig 3). Correlation is "very strong" (tau-c = (.81)[25], p −1.47 e-06, CI=[.66-.96] $a \leqq .01$). Data pair ties limit getting exact p-values; *a* is from Table 1 in Bonett and Wright (2000). Since 5-rank scoring worked well earlier with EC (Fig 5), Fig 7 follows suit with these two common choices. Tau-c is (.72)[25], $p \approx 8$ e-06, CI=[.58-.88] $a \leqq .05$.

Discussion

Rank Order Scales and Uncertainty Metrics

Choice of an uncertainty metric—an evaluation measure—affects how scorings are assessed. One should select a ranking scale in conjunction with the desired metric (or visa versa): For example, when diminished scoring resolution is tolerable, *4-rank/CC* (Fig 3) is a very good pairing. On the other hand, the combination *5-rank/EC* (Fig 5) makes scoring ranks look especially "linear" vis-a-vis uncertainty. Shifting to (scatterplot) comparisons of art images via their median scores and uncertainties, metric EC rates well-enough, but it is dominated by CC. H, a popular but non-linear metric (Zhao and Zhu, 2013; Wang, et al., 2023), gives a less attractive plot of the images. (Jost (2006) suggests H is not always appropriate.) Test before deploying any score/metric combination in actual trials.

The uncertainty of images at higher ranks (4-5) may pose a challenge for the usual scatterplot. Many sculpture pieces are so non-representational that their simplified captions could almost be random selections from a dictionary; as participation increases, so too will the associated uncertainties. This may warrant normalizing uncertainties by the number of participants (for metrics CC, EC) or logarithm of that number (metric H).

"Dislikers"

The *Introduction* alludes to tensions between art abstraction and the public (Wilson and Cupchhik, 1992). Surprisingly, although a plurality of our MTurk participants usually *dislike abstraction*, their responses warrant no special treatment. This contrasts with Wang, et al. (2023), where abstraction rankings emerge only after MTurk "dislikers" are identified (their Figure 7). The authors suggest, "aesthetic preferences may be influenced by the level of engagement, and ambiguous artworks generally require higher engagement ... design choices may make a big difference..." Results here support these remarks anecdotally, although Wang, et al. have a considerably more ambitious agenda incorporating several tasks. Our questionnaire/task involves factors that may affect engagement positively:

1. Clear images
2. Responses are viewer impressions
3. Self-pacing (Wang, et al. mention time as a factor)
4. Real artwork (sculptors fashion their works, likable or not, to engage)
5. Pretesting for acceptability (e.g., shortened—to forestall viewer tedium)
6. Serious respondents (via stiff participant requirements for diligence)

In any case, a third of participants have used an optional comment field to describe our task: 'fun;' 'interesting;' 'excellent;' 'made me think;' 'enjoyable;' etc. One participant extolled, “This was a super fun task and well designed. Definitely a nice break from the typical studies.” That said, controlling for MTurk “typical studies” could prove challenging.

Interesting Outliers

Outliers present opportunities to check on task mechanisms. Consider two image outliers in Fig 7. *Maquette for Radio-Announcer* (1922?)—Fig 7, upper center—has familiar elements; megaphones, parts of a folding stool? child's alphabet block, etc. Its median score is a midrange '3' even though *EC = 34.81* indicates considerable caption variety. Do *Maquette's* everyday elements cause participants to score it lower subjectively than its captions indicate? A milder, complementary example is Hare's *Magician's Game* (1944). Gestalt of *Game (*Fig 7, lower right) is a table, desk or chair for 31 of 60 responses (EC is only 12.68). Still, its many mysterious elements may explain a high median score of '5.' Participants' subjective scorings for both works apparently favor elements over gestalt. A revised questionnaire might investigate this by splitting subjective assessments into two factors, elements and gestalt, each scored separately for abstraction.

/Figure 7 hereabout/

Rules

Categories are established via simple sorting. Optional equivalence rules can then aggregate categories signifying the same thing within a context set by a specific artwork. Synonyms present opportunities to merge category tags. Other word relations, such as metonymy, work as well. Consider rule *crown* ≡ *king*. Merging categories (*crown*, freq.:13) and (*king*, freq.: 9) yields (*crown* OR *king*, freq.: 22). An example rule set, for Chamberlain's *Essex* (1960), is *car* ≡ *automobile, debris* ≡ *garbage* ≡ *junk, dump* ≡ *junkyard* ≡ *landfill, heap* ≡ *pile,* and *jumble* ≡ *tangle*. To demonstrate the effect on partitions, Fig 8 plots the 25 works in black after applying (their 25 optional) rule sets (Fig 7 repeats in light red to highlight changes). *EC* for *Essex* drops from 19.69 categories to 12.68. Other artworks are less affected by their tag equivalences. Merging categories will diminish ambiguity, so the scatterplot shifts lower on the y-axis. Rule sets have one considerable drawback—they are much more subjective than simple label sorting.

New Metrics

Klutsis' *Maquette* serves to illustrate a partition's utility and strength in supporting quantitative metrics beyond the three already covered. It is often remarked of abstract artwork that no one agrees on what is represented. However, partition P contains information that can quantify viewers' level of consensus. Let the specific question be: What prospects do gallery patrons A and B have of agreeing on a representation for *Maquette*? Patron A imagines a caption that fits into category *i* of *P*. This event has a chance ≈ f_i/n, where f_i is the frequency for category *i* and *n* denotes number of participants. Patron B also imagines a caption. For A and B to agree, both captions must fall into category *i*—an event chance of $(f_i/n)^2$. Sum $(f_i/n)^2$ over all *i*, since concurrence may occur with any category. *Maquette's* caption partition is *P = (5, 4, 4, 3, 3, 2, 2, 2, 2, 2, 1,..etc.)*. Category labels for P suggest rules *speaker ≡ loudspeaker; sign ≡ road sign; megaphone ≡ horn.* This gives *P′ = (6, 4, 4, 3, 3, 3, 2, 2, 2, 2, 2,1,...).* Summing category chances of A and B agreeing, $(1 \cdot 6^2 + 2 \cdot 4^2 + 3 \cdot 3^2 + 5 \cdot 2^2 + 27 \cdot 1^2)/(60^2) = 0.039$. A and B likely agree about *Maquette* only 4% of the time. Evidently popular opinion on abstract art has some basis.

6. Summary

Comparisons of subjective abstraction scores vis-a-vis more objective "caption-based" uncertainty (ambiguity) measurements show the two are correlated. Rank order scales are quite good—they have served for years in estimating abstraction levels—but their ranks are sensitive to different assessments from distinct uncertainty metrics. Some combinations of order scales and uncertainty metrics may cause ranks to blur, to be statistically indistinguishable. Top metric performers for this study (i) count distinct types of captions assigned to a work, or (ii) adjust type counts by rebalancing for multiple occurrences. Once uncertainties are linked to ranks to "calibrate" the latter, one has a quantitative sense of how ranks compare. This arrangement offers the speed and ease of scoring along with valid numeric estimates unavailable from ranks alone.

Acknowledgments

We thank Robert Anderson and the referees for suggesting improvements.

References

anon. (2017) Amazon Mechanical Turk Tutorial: Getting great survey results from MTurk and GoogleForms. Happenings at MTurk (web tutorial site).

Agley J, Xiao Y, Nolan R, Golzarri-Arroyo L.(2022) Quality control questions on Amazon's Mechanical Turk (MTurk): A randomized trial of impact on the USAUDIT, PHQ-9, and GAD-7. Behav Res Methods. 2022 **54**(2):885-897. doi: 10.3758/s13428-021-01665-8. Epub 2021 Aug 6. PMID: 34357539; PMCID: PMC8344397.

Atkinson, N.(2015) *Craft for a Modern World: The Renwick Gallery Collection,* Smithsonian American Art Museum in assoc. with D. Giles Limited, London.

Brancusi v. United States, 54 Treas. Dec. 428 (Cust. Ct. 1928).

Bonett, D. and Wright, T. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika* **65**(1), 23-28.

Causey, A. (1998) *Sculpture Since 1945.* Oxford University Press, Oxford and NYC.

Cotter, K. N., Silvia, P. J., Bertamini, M., Palumbo, L., & Vartanian, O. (2017). Curve appeal: Exploring individual differences in preference for curved versus angular objects. *i-Perception, 8*(2), Article 2041669517693023. https://doi.org/10.1177/2041669517693023
Curtis, P. (1999) *Sculpture 1900-1945: After Rodin.* Oxford University Press, Oxford and NYC.
Davidson, J. (1986) The painting continuum. *ETC: A Review of General Semantics* **43**(4), 363-367.
Franklin, M., Becklen, R. and Doyle, C. (1995) The influence of titles on how paintings are seen. *LEONARDO* **26**(2), 103-108.
Gortais, B. (2003) Abstraction and art, *Phil. Trans. R. Soc. Lond. B* **358**(2003), 1241–1249. DOI 10.1098/rstb.2003.1309
Hayn-Leichsenring, G. (2017) The ambiguity of artworks—a guideline for empirical aesthetics research with artworks as stimuli. *Frontiers in Psychology* **8**(article 1857), 1-14.
Hughes, R. (1990) *The Shock of the New, Second Edition*. McGraw-Hill, New York.
Jamieson, S. (2004) Likert scales: how to (ab)use them? *Medical Education*, **38**(12), 1217-1218. (doi: 10.1111/j.1365-2929.2004.02012.x)
Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical Review* **106**(4), 620-630.
Jost, L., Entropy and diversity. OIKOS **113**:2 (2006), 363-375.
Krauss, R. (1979) Sculpture in the Expanded Field. *October* **8**(Spring), 30-44.
Loepp, E. and Kelly, J. (2020) Distinction with a difference? An assessment of MTurk worker types. *Research and Politics* **2020** (January-March), 1-8.
Lucie-Smith, E. (1987) *Sculpture Since 1945.* Universe Books, NY, NY.
Lyssenko, N., Redies, C. and Hayn-Leichsenring, G. (2016) Evaluating abstract art: relation between term usage, subjective ratings, image properties and personality traits. *Frontiers in Psychology* **7**(article 973), 1-9.
McGill, R., Tukey, J.W., and Larsen, W.A. (1978) Variations of Box plots. *American Statistician,* **32**(1)*, 12-16.*
Moszynska, A. (2013) *Sculpture Now*, Thames & Hudson Inc, New York, NY.
Ocvirk, O., Stinson, R., Wigg, P., Bone, R., Cayton, D. (2006). *Art Fundamentals, 10th Edition,* McGraw Hill, NY, NY, p.13.
Pelowski, M., Gerger, G., Chetouani, Y., Markey, P. S., & Leder, H. (2017). But is it really art? The classification of images as "Art"/"Not Art" and correlation with appraisal and viewer interpersonal differences. *Frontiers in Psychology, 8,* Article 1729. https://doi.org/10.3389/fpsyg.2017.01729
Pihko, E., Vitanen, A., Saarinen, V-M., Pannach, S., Hirvenkari, L., Tossavainen, T., Haapala, A. and Hari, R. (2011) Experiencing art: the influence of expertise and painting abstraction level. *Frontiers in Human Neuroscience* **5**(article 94), 1-10.
Read, H. (1964) *Modern Sculpture: A Concise History*, Thames & Hudson Inc, New York, NY.

Shannon, C. and Weaver, W. (1949) *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, IL.
Schapiro, M. (2011) *Modern Art: 19th and 20th Centuries (selected papers),* George Braziller, New York, NY.
Sidhu, D., McDougall, K., Jalava, S., Bodner, G. (2018). Prediction of beauty and liking ratings for abstract and representational paintings using subjective and objective measures. *PLoS One* **13**(7):e0200431, 1-15.
Takahashi, S. (1995). Aesthetic properties of pictorial perception. *Psychological Review, 102*(4), 671–683.

Tormey, J. and Tormey, A. (1983). Art and ambiguity. *LEONARDO* **16**(3), 183-187.

Tucker, W. (1974) *The Language of Sculpture*, Thames & Hudson Ltd., London.
Uusitalo, L., Simola, J. and Kuisma, J. (2012). Consumer perception of abstract and representational visual art. *International Journal of Arts Management* **15**(1), 30-41.
Venables, W.N., Smith, D. M. and the R Core Team (2016). *An Introduction to R*, https://www.r-project.org.

Wang, X., Bylinskii, Z., Hertzmann, A. and Pepperell, R. A computational approach to studying aesthetic judgments of ambiguous artworks. *Psychology of Aesthetics, Creativity, and the Arts* (June 1, 2023), *pages not given*. Advance online publication: https://doi.org/10.1037/aca0000579

Wilson, A.S. and Cupchik, G.C. (1992) The evaluation of high art and popular art by naïve and experienced viewers. *Visual Arts Research* **18**(1)[35], 1-14.
Wilson, A. (2012). Multistable perception of art-science imagery. *LEONARDO* **45**(2), 156-164.
Zhang, K., Harrell, S. and Ji, X. (2012). Computational aesthetics: on the complexity of computer-generated paintings. *LEONARDO* **45**(3), 243-248.
Zhao, M. and Zhu, S-C. (2013). Abstract painting with interactive control of perceptual entropy. *ACM Trans. On Applied Perception* **10**(1), 5:1-5:21.
Zimmer, R. (2003). Abstraction in art with implications for perception. Phil. Trans. R. Soc. Lond. G **358**(2003), 1285-1291.

Glossary

CC, category count

Metric CC(P) counts caption varieties (categories) of an image's partition P. *Sometimes called a species count.*

EC, *equalized category count*,

is an uncertainty metric $2^{H(P)}$ that counts categories of an idealized partition P* having all frequencies equal and H(P*) = H(P).

$\underline{EC}$, *median EC*

H, ensemble (or Shannon) entropy

is an information-theoretic metric; it characterizes disorder in P in terms of **bi**nary digi**ts** (**bits**) needed, on average, to identify a partition category (Shannon and Weaver, 1949).

M, a metric

M(P) measures ambiguity/uncertainty for *P* (below). *M ≡ CC, H or EC*

P, partition

A list of frequencies, each of which indicates occurrences of one caption variety for an artwork.

Appendix A: Classifying Captions by Spellings

Image captions are reduced to character string labels (tags) that denote partition categories. Optional semantic constraints are applied after this stage.

For each caption:

1. Ignore upper and lower case.
2. Identify the caption subject, e.g., "several white geese alighting on a pond" concerns "several white geese."
3. Reduce the subject to a simple subject, if this is possible (otherwise, see #3 and #4, below). For the above, this is 'geese.'
    - Treat singular and plural forms as equal, so for classification, goose ≡ geese.
    - Determine what *is* a single word via the *Merriam-Webster (MW)* on-line dictionary/thesaurus. "Holy man"—having no *MW* entry—reduces to 'man.' In contrast, 'wise man' is a *MW* entry and will be recorded as 'wise man,' a category separate from 'man.' 'Junk yard' can be spell-corrected to 'junkyard' since the latter is an entry in *MW*. Note that 'Canada geese' is also a *MW* entry and is classified separately from 'goose.'
    - Other than singular/plural forms and spelling corrections, there are no word equivalents. Do not use a thesaurus; for this phase, 'pupil' is not the same as 'student.' Each defines a separate category. This may seem crude but (a) it works and (b) it supports having label/tag equivalences such as junk ≡ trash expressed in a separate, later phase.
    - If the caption is an imperative, e.g., "Spread the news," the simple subject is an implied 'you' that tells little. Therefore, an imperative caption is treated as an exclamation. See #4, below.
4. Simplify compound subjects but keep them compound. The subject "green grapes and red apples" becomes "apple and grape." (Note alphabetical re-ordering for sorting.)
5. Take literals, imperatives, expletives, expressions, tags and exclamations just as they are. 'Meow meow' thus becomes a classification tag, as does '*!**!!%##.' Different names of the same thing are distinct categories. For example, Jenga Building and

Jenga Tower denote a structure at 56 Leonard St., NYC, but this phase does not recognize this fact.

6. Add a new rule to this list to handle an unforeseen case, but then apply this rule to those captions already classified that it affects.

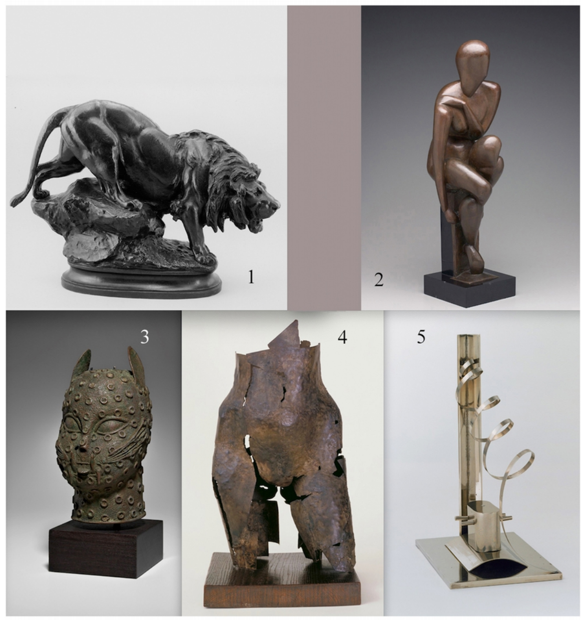
One may contest *MiriamWebster* via other sources, but doing so misses the point: *MW* is the selected arbiter, a standard against which others can re-check classifications.

Appendix B: Sculptural Works and their Image Sources *(see page 21)*

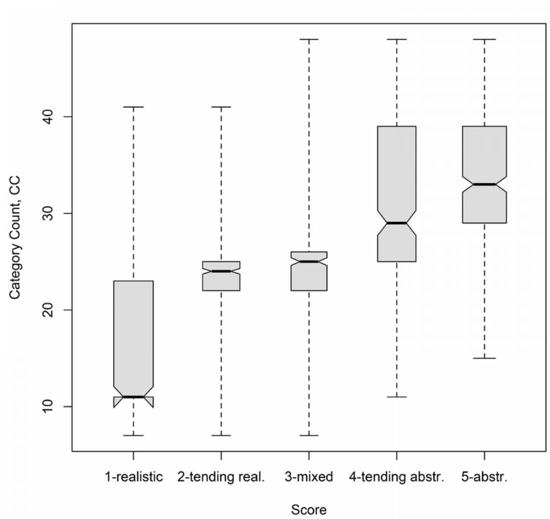
Figure Captions

Figure 1. Sculptures of Five Abstraction Ranks
Figure 2. Category Count versus Score (1500 responses)
Figure 3. Category Count versus Score, 2-3 Merged (1500 responses)
Figure 4. Entropy (base 2) versus Score (1500 responses)
Figure 5. Equalized Category Count versus Score (1500 responses)
Figure 6. Scatterplot of 25 Images: CC vs. Median Score
Figure 7. Images : EC vs. Median Score
Figure 8. Images : EC with Rules

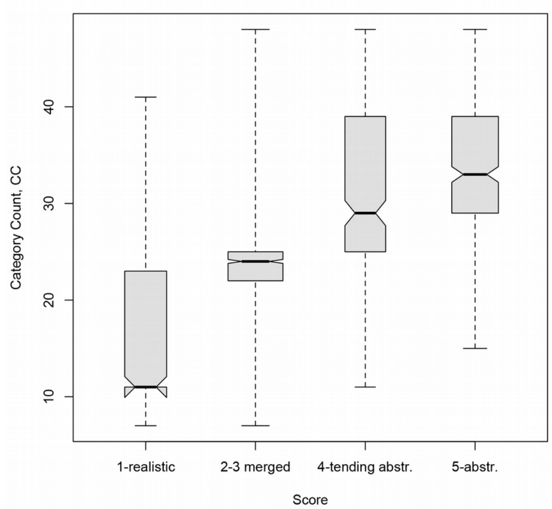
<Figure 1>



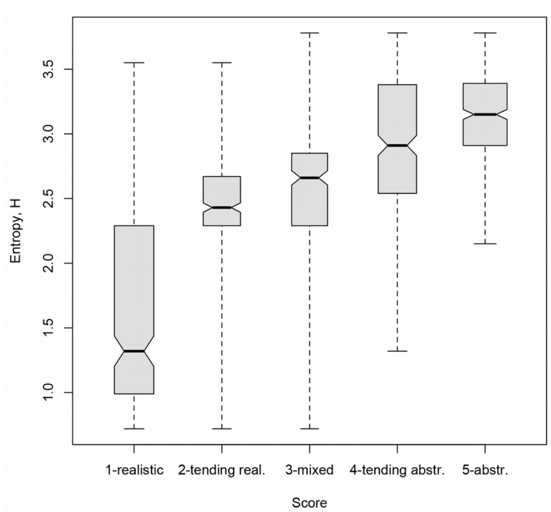
< Figure 2 >



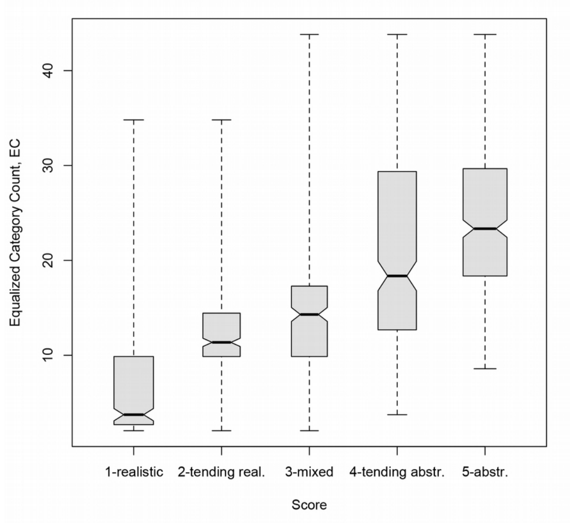
<figure 3>



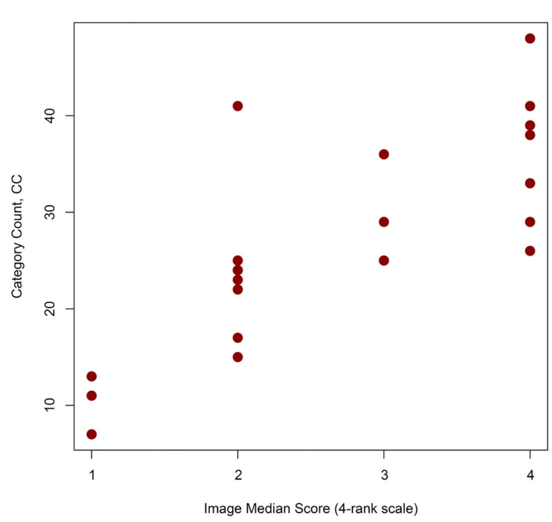
<figure 4>



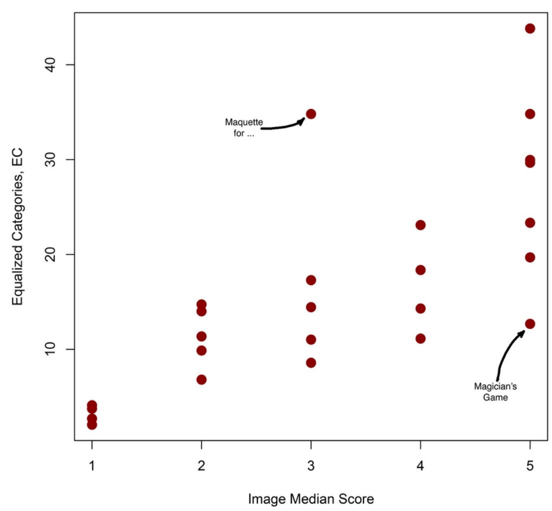
<figure 5>



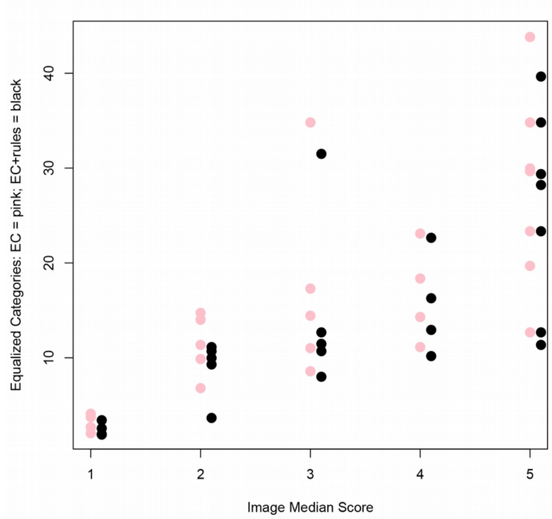
<FIGURE 6>



<FIGURE 7>



<FIGURE 8>



Sculptor	Year	Title	Median Rank	Source	Copyright Status	Alternate Access or View
Anon.--Mayan Culture	600-800	*Ballplayer*	*2*	https://artmuseum.princeton.edu/collections/objects/37593	Fair use for individuals; details at bottom of displayed image	
Bonheur, Rosa	1880	*Lion*	*1**	https://www.artic.edu/artworks/70989/lion	CC0 Public Domain Designation	
Edo, anon.	20th cent.	*Head of Leopard*	*3**	https://artgallery.yale.edu/collections/objects/84391	Copyright undetermined	
Duchamp-Villon, Raymond	1914	*Seated Woman*	*2**	https://artgallery.yale.edu/collections/objects/48055 (image 1)	CC0 Public Domain Designation	
Moholy-Nagy, Laszlo	1921	*Nickel Construction*	*5**	https://www.wikiart.org/en/laszlo-moholy-nagy/nickel-construction-1921	Public domain	
attr. to Klutsis, Gustav	1922	*Maquette for Radio Announcer*	*3*	https://www.pinterest.com/pin/454441418623414955/	JStor/ARTStor Terms and Conditions	https://www.jstor.org/stable/community.15655722
Robus, Hugo	1933	*Girl Washing her Hair*	*4*	https://www.jstor.org/stable/community.14631010	JStor/ARTStor Terms and Conditions	https://www.metmuseum.org/art/collection/search/486885 ; metal version. JStor image is similar, but shows marble variant.
Davis, Emma	1934	*Cosmic Presence*	*3*	https://www.jstor.org/stable/community.13578812	JStor/ARTStor Terms and Conditions	MoMA site mentions work but omits any image.
Gonzales, Julio	1936	*Torso*	*4**	https://www.wikiart.org/en/julio-gonzalez/torso-1936	Public domain	
Martins, Maria	1941	*Christ*	*2*	https://www.moma.org/collection/works/81589?artist_id=3767&page=1&sov_referrer=artist	Licensing via Art Resource (N.America) or Scala Archives (elsewhere)	
Esherick, Wharton	1942	*Reverence*	*3*	https://www.jstor.org/stable/community.14633168	JStor/ARTStor Terms and Conditions	https://whartonesherickmuseum.org/sculpting-in-remembrance-a-grave-marker-for-sherwood-anderson/ (penult photo, of wood version, is similar to that from Jstor except for child viewer)
Hare, David	1944	*Magician's Game*	*5*	https://www.wikiart.org/en/david-hare/magician-s-game-1944	© David Hare: Fair Use	https://www.jstor.org/stable/community.14560870
Martins, Maria	1945	*The Impossible*	*5*	https://www.jstor.org/stable/community.14382446	JStor/ARTStor Terms and Conditions	https://artsandculture.google.com/asset/the-impossible-maria-martins/pgHUga8PqrukaQ?hl=en (Martins' work from different angle)
Howard, R.B.	1950	*The Miscreant*	*5*	https://www.jstor.org/stable/community.14620338	© 2007 Artists Rights Society (ARS), New York	
Chamberlain, Robert	1960	*Essex*	*5*	https://www.jstor.org/stable/community.13588788	2009 John Chamberlain / Artists Rights Society (ARS), New York	Similar image: https://www.tumblr.com/archiveofaffinities/106911807851/john-chamberlain-essex-1960
Tucker, William	1967	*untitled*	*4*	https://www.moma.org/collection/works/82027?artist_id=5963&page=1&sov_referrer=artist	JStor/ARTStor Terms and Conditions	https://www.jstor.org/stable/community.14552922
DeMaria, Walter	1967	*Cage II*	*4*	https://www.wikiart.org/en/walter-de-maria/cage-ii-1965	© De Maria Fair Use	
Ferrara, Jackie	1979	*A200AJUT*	*3*	https://www.moma.org/collection/works/81985	Licensing via Art Resource (N.America) or Scala Archives (elsewhere)	https://www.jstor.org/stable/community.15661706
Minujin, Marta	1981	*Falling Venus*	*2*	https://www.jstor.org/stable/community.14379500	JStor/ARTStor Terms and Conditions	Similar image: https://libreriaelastillero.com/2914-large_default/marta-minujin-venus-de-milo-1980.jpg
Shea, Judith	1983	*Crawl*	*2*	https://www.moma.org/collection/works/100328 (image 1)	JStor/ARTStor Terms and Conditions	https://www.jstor.org/stable/community.15653716
Deacon, Richard	1984	*Fallling on Deaf Ears*	*5*	https://www.moma.org/collection/works/81378 (image 2)	Licensing via Art Resource (N.America) or Scala Archives (elsewhere)	https://library.artstor.org/asset/AMOMA_10312310107
Spiering, Ken	1989	*Childhood Express*	*1*	https://www.jstor.org/stable/community.14743120	JStor/ARTStor Terms and Conditions	Similar image: https://foursquare.com/v/the-childhood-express/4bca4bd00687ef3b0eccdbcc?openPhotoId=5bbbbc2673fe25002cc21864
Gober, Robert	1989	*Cat Litter*	*1*	https://www.moma.org/collection/works/81486	Licensing via Art Resource (N.America) or Scala Archives (elsewhere)	https://www.jstor.org/stable/community.14558513
Benedit, Luis	1990	*Sugar Ranch Hous*	*1*	/https://www.jstor.org/stable/community.14371213.	JStor/ARTStor Terms and Conditions	
Ferrari, Leon	2006	*Huesos (bones)*	*5*	https://www.jstor.org/stable/community.14185514	JStor/ARTStor Terms and Conditions	Similar image: https://www.moma.org/audio/playlist/233/2997