

Supplementary methods and evidence

Measuring Value in AI Generated Visual Art

Liyuantan, lichnagtan • 20 October 2026

S1 Corpus and record structure

The accompanying evidence.json and evidence.csv contain 12 reports, 25 study units and 46 outcome-family records. The report is a publication; the study unit is an eligible main experiment or study within that publication. An outcome-family record identifies content and response procedures. Multiple items can share a family. A composite can contribute to more than one family when its items combine different content. Records are therefore not independent samples or a count of separate statistical endpoints.

The report identifiers are R01, R02, R03, R04, R05, R06, R07, R09, R12, R15, R16 and R17. Gaps reflect source-access work, not missing included reports. Study identifiers preserve original study or experiment numbers; Lab and Online distinguish the visual-art samples in Lane et al. Outcome identifiers append O1, O2 and so on within a study. CSV and JSON represent the same records. The CSV uses UTF-8 with a byte-order mark.

S2 Decision rules and uncertainty

Read the recorded response before assigning a family. If economic wording is answered on an ordinal or agreement scale without currency, classify it as a noncurrency economic rating. If the answer is monetary, identify whose payment is being judged. The respondent's own hypothetical payment belongs to personal stated WTP; an anticipated or recommended selling price belongs to sale-price judgment, with the predictive or normative wording preserved. An acquisition likelihood or relative buying preference belongs to purchase intention. A documented auction with payment and allocation consequences belongs to consequential bidding.

Question content and aggregation must both be retained. Chamberlain et al.'s Study 2 includes economic wording within a mean aesthetic-value composite. Its two family records identify this mixed content; they do not imply separate outcome analyses. Aesthetic appraisal includes liking, beauty, creativity and related judgments for this descriptive map, without asserting one common latent scale.

Financial consequence concerns the evaluative response itself. Payment for completing a survey does not turn a rating into a purchase. For the retained auctions, one randomly selected round determines payment and allocation. A bid is the submitted offer; the second-highest bid determines the winner's price. The extraction describes reported implementation, rather than auditing individual transactions.

Actual source, attributed source, viewing setting and acquisition object are independent descriptions. A human painting assigned an AI label remains human-made in the actual-source field. An art-fair setting can contain a hypothetical rating task, and an online study can deliver a physical print. Where the source is incomplete, the relevant detail is marked unverified.

S3 Search record and access decisions

Search and follow-up retrieval were conducted on 20 October 2026. Targeted web discovery, reference checking and exact-title retrieval supported a purposive methodological corpus. The

following retained queries document the search approach; they are not a complete export of every preliminary query, and no screening total is reconstructed.

- "AI-generated art" "willingness to pay" experiment
- "AI art" "purchase intention" human experiment
- site.pec.ac.uk "A Silver Lining" art
- "A Silver Lining" "Lane" "PEC"
- "Creativity and aesthetic evaluation of AI-generated artworks" psychology methods
- "Can Artificial Intelligence Exert Effort" 2026 DOI
- "Art as revelation" Depecik pdf
- "Framing the Future" "Jin" "Tao" pdf
- "The creativity of artificial intelligence" "Prissé" pdf
- "Prissé" "Deng" "creativity" auction pdf
- "Neef" "TS4V3" codebook
- "hypothetical bias" "willingness to pay" "meta-analysis" Schmidt Bijmolt 2020
- Flake Fried 2020 measurement schmeasurement questionable measurement practices DOI full text

The Neef article links its public OSF project at <https://doi.org/10.17605/OSF.IO/TS4V3> . The codebook was retrieved from <https://osf.io/download/namgu/> . In sheet Codebook, rows 12–20 for AA01_01 describe attractiveness and rows 487–495 for KI21_01 describe buying intention. Codes 1–7 range from strong disagreement to strong agreement; -1 denotes inability to judge and -9 denotes no answer. The participant data file was not needed for this extraction.

The Prissé SSRN record was located at <https://doi.org/10.2139/ssrn.5284780> . Direct full-text retrieval remained unavailable. Its public institutional project description and the final publication record contain different descriptions of the image-generation system, so the earlier project page was not used as a substitute for the final methods.

Follow-up retrieval located Chamberlain et al.'s publisher supplements at https://supp.apa.org/psycarticles/supplemental/aca0000136/aca0000136_supp.html . Appendix 3 was visually inspected because the questionnaire is embedded as a graphic in the Word file. Appendix 4 supplied the retained factor-loading items. The full follow-up query list and access decisions are in revision-search-log.json. No new report or study unit was added.

Table S1 Material source access decisions

Candidate	Access and inclusion	Basis
R01–R07	Core texts retrieved	Publisher or institutional texts; R06 supplement, R04 codebook and R07 publisher Appendices 3–4 inspected.
R08 Jin and Tao 2025	Detailed coding withheld	10.1002/mar.22242. Abstract accessible; complete methods not retrieved.
R09 Lane et al. 2026	Core full text retrieved	Creative PEC PDF. SSRN request failed; public centre version obtained. One report across versions.
R10 Depecik et al. 2026	Detailed coding withheld	10.1016/j.ijresmar.2026.09.002. Publisher preview available; institutional PDF staff-restricted. Author sharing links found but redirects failed.

Candidate	Access and inclusion	Basis
R11 Hu et al. 2026	Detailed coding withheld	10.1016/j.jretconser.2026.104843. Publisher abstract/snippets located; complete methods not retrieved.
R12 Theirs et al. 2026	Core methods read in HTML	10.1525/collabra.164122. Publisher HTML accessible through web retrieval; direct file request failed.
R13 Hong and Curran 2019	Detailed coding withheld	10.1145/3326337. Publisher file request failed; no complete legitimate text retrieved in this workflow.
R14 Gangadharbatla 2022	Detailed coding withheld	10.1177/0276237421994697. Publisher file request failed; no complete text used for coding.
R15 Chiarella et al. 2022	Core full text retrieved	Official Sapienza university PDF obtained with normal system web client.
R16 Hall and Schofield 2025	Core full text retrieved	Publisher PDF; small convenience sample and undisclosed actual source noted in appraisal.
R17 Cunningham et al. 2025	Core full text retrieved	Frontiers publisher PDF.
R18 Messer 2024	Detailed coding withheld	10.1016/j.chbah.2024.100056. Publisher method excerpts located during follow-up retrieval; complete source and supplements not retrieved or fully extracted.
Prissé et al. 2026	Detailed coding withheld	10.1016/j.socec.2026.102528. Auction described in abstract; payment implementation not verified.

Oksanen et al. and Bianchi et al. provide art-review context. Flake and Fried provide a measurement-transparency framework; Schmidt and Bijmolt provide the broader distinction between hypothetical and consequential WTP elicitation. These four publications are references for interpretation, not primary AI-art study units. Music and creative-writing studies, startup tasks, advertising design and technical image benchmarks do not contribute to the counts.

S4 Study interpretation notes

Table S2 Interpretation checks

Report	Measurement and design considerations
Bellaiche et al. (2023)	R01-S1: Only failed label-check trials removed; 113 of 4470. A label effect on fixed image content. R01-S2: Time estimate is duration, not money. Reuses stimuli from Study 1; 114 of 4440 trials excluded.
Horton et al. (2023)	E1–E2 use rating scales; E3–E6 ask both recommended seller price and personal payment under fictional gallery information. No purchase mechanism. Experiment 5 has inconsistent reported N. Shared stimuli and repeated ratings are not independent samples.
Gu and Li (2022)	R03-S1: Human attribution includes famous-artist description. A non-significant source contrast is not an equivalence test. R03-S2: Expertise groups differ in recruitment context; expertise was not randomly assigned.

Report	Measurement and design considerations
Neef et al. (2025; online 2024)	R04: Joint versus separate source evaluation. Outcome endpoints verified in the public codebook: 1–7 agreement; -1 cannot judge and -9 unanswered are special codes.
Millet et al. (2023)	R05-S2: Relative slider position is not an independent rating of each image. R05-S3: Measured anthropocentric creativity beliefs are not a randomized mechanism. R05-S4: Explicitly hypothetical; no price, budget, no-buy option or checkout. Study 1 is music and excluded.
Magni et al. (2024; online 2023)	R06-S1: Composite content differs from liking and from the later originality-based creativity scale. R06-S4: Authors call the supplementary outcome WTP; its question asks predicted selling price, not the participant's own maximum payment. Effort check uses a different response format from creativity.
Chamberlain et al. (2018)	R07-S1: Subjective source classification is endogenous to image appearance; not a randomized human/AI label experiment. R07-S2: Different venues, dates and sample composition across conditions. The 11-item aesthetic-value composite includes two five-category economic agreement items, confirmed by publisher Appendices 3 and 4. Three reported condition sizes sum to 349; Table 4 instead gives N=353, and some other analyses use smaller denominators. These discrepancies are retained without reconstructing participant flow.
Lane et al. (2026)	R09-Lab: One paid round per four-person group; printed objects distributed. Session-level assignment and selected image pools constrain inference. R09-Online: Print mailed; cost deducted from £12 bonus. Raw means £5.68 human and £5.29 AI, $p=.191$; fully adjusted Tobit coefficient for AI $-\pounds 0.82$, $p<.05$. Latent-index coefficient is not an observed mean difference.
Theirs et al. (2026)	R12-S1: Time manipulation differs from source manipulation. Ratings are not currency amounts. R12-S2: Recruitment paragraph gives 240 initial and 209 final, with exclusion categories 5 and 31; overlap is not explained there. Final N preserved without reconstructing a participant flow.
Chiarella et al. (2022)	R15: 67 self-described art experts, 43 nonexperts. Art-fair context does not imply purchases; physiological outcomes are not economic values.
Hall and Schofield (2025)	R16: Small convenience sample; no manipulated source label and no transaction procedure. Three pairs have higher human means, one equal, one higher AI mean; no general label premium follows.
Cunningham et al. (2025)	R17: Questionnaire assesses the collection, not a separate response after each image. Eye movements do not directly measure willingness to buy.

S5 Coverage checks

Counts were recalculated after removing each report in turn. The final row also removes R07, the historical computer/robot-art report, and R16, the 27-person five-pair valuation study. This additional exclusion tests whether the measurement map depends on these two boundary cases; it is not a quality-score threshold. All checks were developed during revision and are exploratory.

Table S3 Exploratory coverage after report exclusions

Omitted	Studies	AE	ER	SP	WTP	PI	AB
None	25	23	7	5	5	4	2

Omitted	Studies	AE	ER	SP	WTP	PI	AB
R01	23	21	5	5	5	4	2
R02	19	17	5	1	1	4	2
R03	23	21	7	5	5	2	2
R04	24	22	7	5	5	3	2
R05	22	20	7	5	5	3	2
R06	23	21	7	4	5	4	2
R07	23	21	6	5	5	4	2
R09	23	23	7	5	5	4	0
R12	23	21	5	5	5	4	2
R15	24	22	7	5	5	4	2
R16	24	22	7	5	4	4	2
R17	24	22	7	5	5	4	2
R07, R16	22	20	6	5	4	4	2

AE denotes aesthetic or artistic evaluation; ER, noncurrency economic rating; SP, sale-price judgment; WTP, personal stated monetary WTP; PI, purchase intention or relative buying preference; AB, consequential auction bid. Columns count study units and can overlap. Removing R02 leaves one SP unit and one WTP unit; removing R09 removes AB entirely. All six families remain after the combined R07/R16 exclusion. These are coverage results, not sensitivity estimates for a pooled preference effect.

S6 Coding clarifications and remaining uncertainties

Neef et al.'s response categories were established from the linked codebook. Chamberlain et al.'s mixed composite contributes both aesthetic and economic-content records. The publisher questionnaire shows that its drawing statements use five agreement categories from strong disagreement to strong agreement. The semantic-differential items concern the robots. The questionnaire does not print numeric codes for the agreement options.

Horton et al.'s seller question is normative; Magni et al.'s is predictive. Both belong to the descriptive sale-price-judgment family, with their wording retained in the records. The machine-readable key `estimated_sale_price` covers both forms for compatibility. Horton's six source-attribution descriptions distinguish the condition structure of each experiment. Source labels are recorded separately from the actual provenance of the images.

Chamberlain et al.'s reported condition sizes sum to 349, while Table 4 gives N=353 and some analyses use smaller denominators. Horton et al.'s Experiment 5 reports 709 in Results and 710 in Methods. Theirs et al.'s Study 2 reports 240 recruited and 209 retained alongside exclusion counts of five and 31 without clarifying overlap in that paragraph. These discrepancies remain explicit.

Hall and Schofield report dollar responses without establishing the complete elicitation bounds or currency jurisdiction in the inspected methods. The individual provenance of the two fixed images used in Horton et al.'s Experiments 3 and 4 was not independently resolved here. Millet et al.'s Study 4 uses adopted poster stimuli whose exact production provenance was not independently established. The unavailable detailed transaction methods of Depecik et al. and Prissé et al. remain priorities for an expanded review.

S7 Files and reproduction

evidence.json preserves the report–study–outcome hierarchy. evidence.csv provides the same entries as rows. evidence-codebook.txt defines the fields and permitted categories. corpus-counts.json contains both study- and report-level category counts. coverage-checks.json contains each exclusion scenario.

Run reproduce_counts.py in the directory containing these files. Python's standard library is sufficient. The script checks identifier uniqueness, record completeness, agreement between CSV and JSON, and consistency of the saved counts and coverage scenarios. It writes counts-reproduced.json. It does not reproduce the original studies' statistical analyses or establish the correctness of an extraction against a source.

The source-location and appraisal fields support substantive verification. Both authors have read the manuscript. No independent human extraction agreement or independent audit of all extraction records is claimed. Source PDFs and the external codebook are cited rather than redistributed in the supplementary archive.

S8 Targeted claim audit

The companion claim-audit.json contains eight checks linking a substantive claim to a source, location, classification decision and remaining uncertainty. They cover Horton's scaled and monetary questions, Magni's selling-price prediction, Chamberlain's composite and response categories, Lane's settlement mechanism and model interpretation, and the missing transaction reports. The checks were conducted within the same AI-assisted workflow. They do not provide independent extraction agreement.

Figure 1 is an original diagram of the review's coding rules. It is supplied as a separate high-resolution image and vector PDF. It is not reproduced from a source questionnaire and does not validate the measurement families. The worked examples point to already included reports; they introduce no additional observations.