cyberwitchery papers / CW:002

Teaching Personal Taste to Silicate

A drawing machine for tacit thought

Veit Heller

cyberwitchery lab

September 2026

case study · version 1.0

"The old world is dying, and the new world struggles to be born: now is the time of monsters."
— Slavoj Žižek, after Antonio Gramsci[1]

abstract

Algorithmic and generative art move creation away from the singular gesture and into a field of possible works. When the machine can create without tiring, abundance shifts the artist's labour toward curation, refusal, and guidance. Taste, once a background skill, comes to occupy much of the room left for human agency.

This paper presents a drawing instrument that remembers one artist's quick, tacit judgments—*love*, *meh*, and *hate*—through an authored vocabulary of 222 decisions. Across 2,196 ratings, reconstructed historical models recover 22.4% of the variation in the judgments that follow, enough to change what appears next, but not enough to settle the matter. Reading that partial success alongside the compositions the vocabulary cannot explain, we examine what happens when taste is given a technical body, and why the machine's useful misunderstandings may matter as much as its agreement.

keywords generative art; preference learning; interactive evolution; human–computer co-creation; computational aesthetics

claim in one line

The machine does not need to know what beauty is, but it can learn to emulate feeling.

[1] This wording appears in a 2010 paraphrase of Gramsci (Žižek 2010, p. 95). In the standard English translation, Gramsci's passage speaks instead of the "morbid symptoms" that appear in the interregnum (Gramsci 1971, p. 276).

1 The room occupied by taste

Generative art moves creation away from the singular chosen gesture into a field of possible gestures. The artist builds a system, gives it materials and laws, then watches it produce. Authorship does not disappear in this plane, it spreads into constraints, parameters, seeds, refusals, and the decision that this image speaks while that one does not. Questions of agency have followed generative art for as long as the field has had a name (McCormack, Bown, et al. 2014); abundance makes them impossible to keep theoretical.

When the machine can make without tiring, making is no longer the only scarce act. Attention becomes scarce. Selection becomes scarce. The artist spends less time asking how to produce another image and more time asking whether the image is worth keeping, whether its failure is inherent to the mechanism, whether something in it should be allowed to reproduce. Taste and curation become the forefront of the creative act.

Taste is a troublesome word. It sounds exacting, almost scientific or at the very least empirical, when in practice it is private, contextual, and often inexpressible through language. We know before we can explain. We change our minds. A form loved yesterday becomes exhausted through repetition, an ugly accident teaches us to see things differently. In this paper, I treat taste not as a universal score hidden inside an image. It is the wake left by an artist moving through one changing space of unexplained decisions.

Much machine learning for image aesthetics asks a different question. NIMA predicts distributions of aggregated ratings of photographic quality (Talebi and Milanfar 2018), personalized ranking admits that an individual's order may differ from the crowd (Lv et al. 2018). Preference learning gives us ways to learn from choices rather than declarations (Fürnkranz and Hüllermeier 2010), while interactive evolutionary computation has long let human judgment become the fitness function of a search (Takagi 2001). CLIP-guided drawing, in turn, steers an image toward the semantic gravity of a prompt (Frans et al. 2022). Closer to this work, McCormack and Lomas learn one artist's aesthetic rankings from both rendered images and generative parameters, then use those predictions to guide further exploration (McCormack and Lomas 2020).

The instrument here presented takes a deliberately narrower, more legible route: can a machine learn enough of a single artist's recurring, half-spoken judgments to become a useful accomplice in the creative act? It does not inspect pixels or consult a population. It watches the choices made inside a procedural drawing system—palette, material, density, placement, gesture—and remembers which combinations were met with *love*, *meh*, or *hate*. That memory then shapes the next field of images. The crudeness is deliberate: a finer scale invites thought and rationalization, while these three quick judgments are meant to arrive before explanation. Love, hate, and ambivalence are expressions of the subconscious if wielded correctly.

This is not an attempt to solve aesthetics, it is an attempt to build a studio instrument that can be taught, and to examine what happens when the most liminal part of the creative act is given a technical body. The central question is what is representable, and whether the parts that are not are too crucial to be left out.

2 A drawing machine capable of remembering

The workflow is modeled as a closed loop between human judgment and machine generation. The implementation uses a compact recommendation representation and expands it into a full, reproducible render program before drawing.

It is deliberately simple, even crude. The mechanism was made to aid rather than distract from the creation, and its field of authority is narrow but powerful.

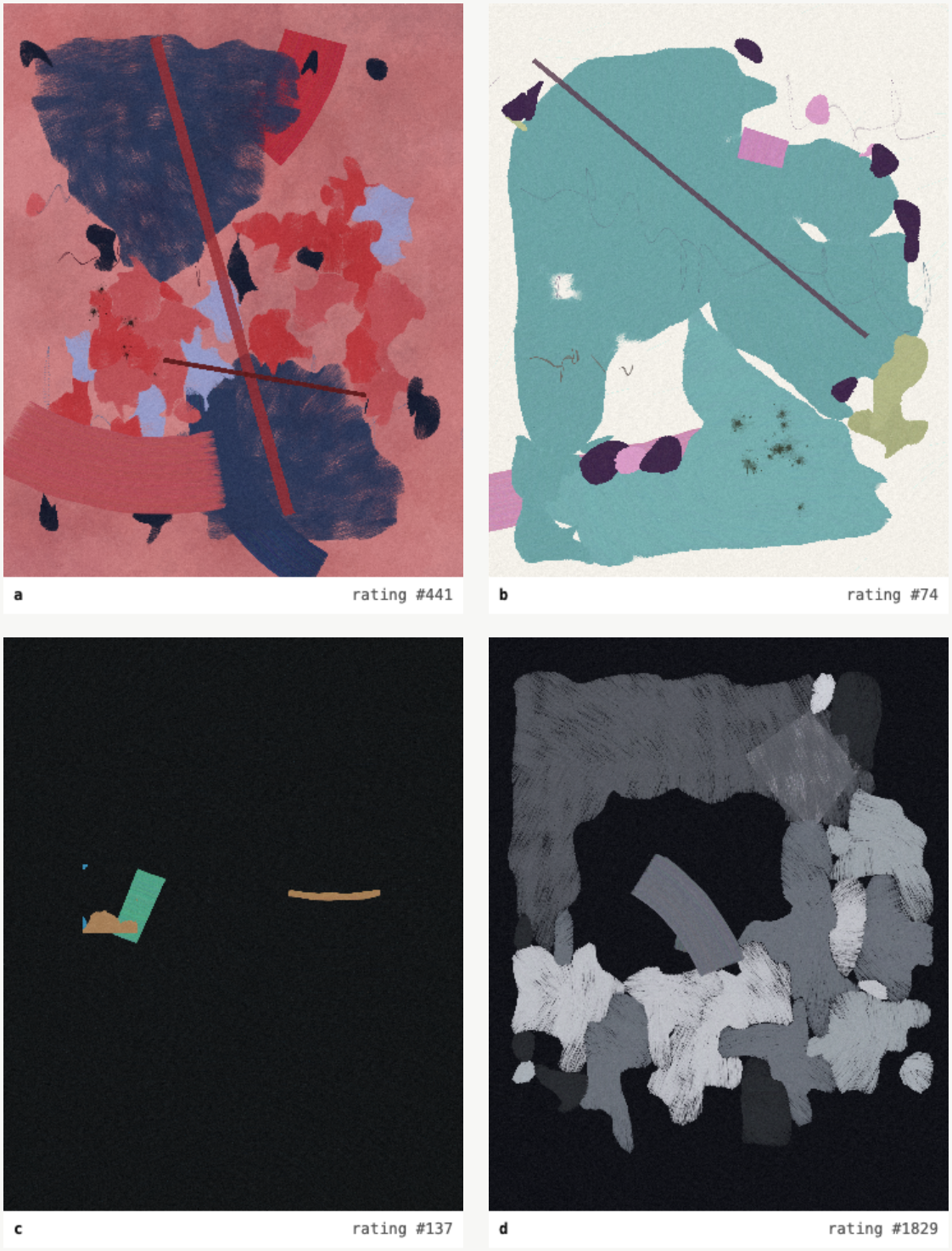

Figure 1. Four stored recipes rendered with the current sketchbook code. The upper pair received *love* ratings and high fitted scores; the lower pair received *hate* ratings and low fitted scores. The panels are labelled a–d in their corners. They are deliberately selected contrasts, not estimates of an average loved or hated image.

2.1 The visual substrate

The creature underneath this experiment is a browser-based p5.js renderer. It produces 1600×2000 pixel canvases at pixel density two, but borrows its language from matter: felt-tip fields, pencil hatching, watercolor washes, ink splatter, brush marks, pointillist dust, voids, and sand-like scribbles, algorithms created and collected over years of artistic expression. The computer is fast, but the work is not meant to look frictionless. Its simulated fibres, uneven fields, and awkward edges are a deliberate resistance to the clean infinity that software offers when unchallenged.

This vocabulary belongs to the instrument, not to the learning method. Another renderer could inherit the same feedback loop, but it would bring different nouns, different accidents, and therefore a different theory of what can be noticed. As such, it is an artist's toolbox separate from the learning machine.

A drawing begins as a compact *recipe*: a seed, one of 15 compositional modes, one of 11 palette moods, one of four background families, one of seven texture profiles, a set of traits, and a list of elements placed in space. The recipe is a proposition, not yet an image. Normalization unfolds it into an executable render configuration, synthesizing concrete colours, trait parameters, overlays, and field settings deterministically from the random seed. Only then does the renderer lay down background, overlays, stroke fields, and paper noise. As such, it is a reproducible kind of randomness, the field where most generative art lives.

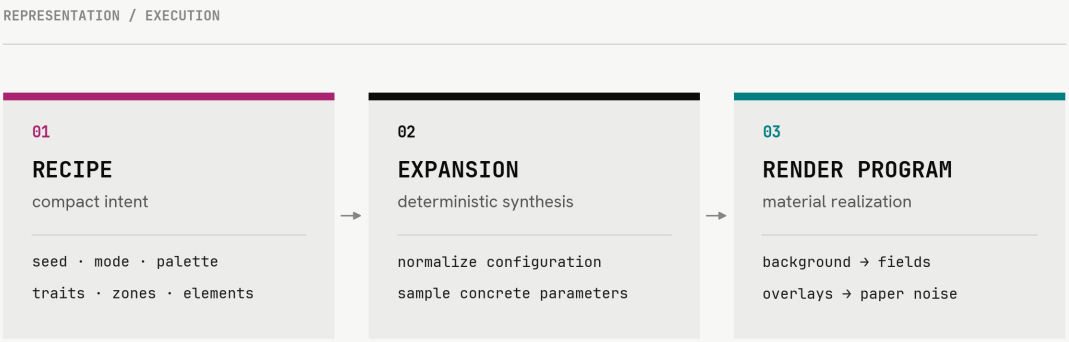

Figure 2. A stored recipe is expanded into an executable visual program. The preference model scores the compact description, while the renderer remains free to realize it through a fresh seed.

This tension between reproducibility and randomness, while not a core proposition of this paper, is still a core feature of the instrument. A fixed seed preserves the original drawing, whereas a new seed keeps the broad artistic proposition while letting its local accidents happen differently. The machine remembers a family resemblance rather than embalming one successful child. Without that variation, the search risks premature convergence, an important failure mode for the system as a whole.

2.2 The feedback cycle

The interface presents images in batches. Each receives a quick *love*, *meh*, or *hate*, mapped to 2.5, 0.5, and −1.5. Their offset lets *meh* reproduce; *hate* is refusal. Rating and

recipe enter an SQLite database. Between batches, the engine rebuilds its model, breeds from prior recipes, and chooses a next group balancing promise with difference. A render becomes a judgment through review, altering the space of future arrivals.

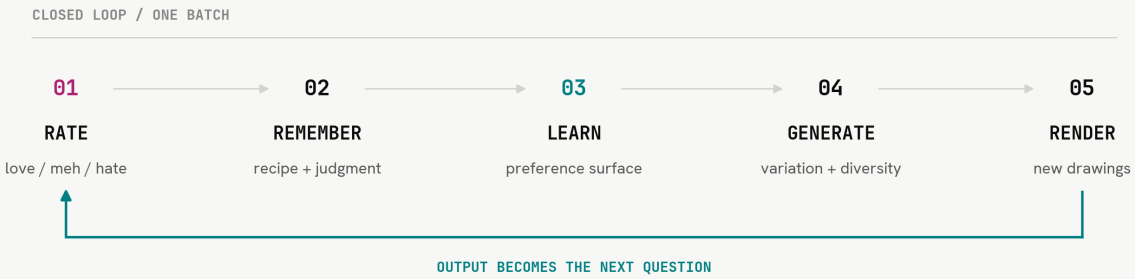

Figure 3. The operational loop. A judgment is both an evaluation of an output and training data for the distribution from which later outputs arrive.

The resulting archive is, by design, not a neutral survey. Later drawings descend, in part, from earlier judgments. The 2,196 observations in the current database are footprints made across a shifting path. Statistically, this means they are not independent samples from a fixed distribution.

Artistically, that is the point. An instrument that learns should change shape as it is played under the player's fingers. The player, hearing the change, also changes how they play.

Figure 1d (lower right) is a useful complication. Its *hate* label was a verdict on the image as a whole, made in the moment, not a claim that all of its constituent choices were bad. The drawing has redeeming material and local qualities, but the composition felt flat and boring. Its parts did not produce enough tension, hierarchy, or movement. The three-way label collapses that ambivalence into a single target. Capturing this kind of felt, interaction-level judgment—when individually promising qualities fail as a composition—is the explicit aim of the instrument, and where actual utility arises.

3 An authored vocabulary for taste

For the machine to remember taste, taste first has to be made addressable. Let c be a recipe and $\phi(c) \in \mathbb{R}^{222}$ its feature vector. The 222 dimensions do not fully describe what the finished picture looks like. They describe systemic decisions.

The largest block crosses twelve trait types with nine spatial zones. A brush stroke in the top left is not the same proposition as a brush stroke in the center. Other features record how many anchors and chips exist, how large they are, how far they spread, which zones they occupy, how the field bends, and which named modes, palettes, backgrounds, and textures were invoked. Of the 222 dimensions, 218 actually vary in the recorded work.

This component of the work is part engineering, part ontology, and presupposes artistic choices by defining a vocabulary. It can encode palette, density, placement, and named gesture. It cannot define a colour harmony that emerges only after rendering, the balance of the whole page, the semantic echo of a mark, the artist's mood, or the lucky curve that suddenly makes everything around it necessary.

block	dimensions	information retained
zone–trait	108	Presence of 12 trait types in each cell of a 3×3 grid.
global trait	13	Count of each trait type and total trait count.
derived trait	17	Coverage, diversity, overlay count, center/edge density, and a contrast signal.
composition/render	43	Element counts, coverage, centroid, spread, zone occupancy, field density, warp, margins, and size ranges.
categorical	41	Orientation, 15 modes, 11 palettes, four backgrounds, and seven texture profiles.
total	**222**	

Table 1. The feature ontology. It records choices available to the generator, not measurements extracted from the final pixels.

Its limitations might make it appear crude, but they produce a vehicle that suits its artist. It is also extensible, and mirrors an artist's palette more than a creative godbox.

A model trained on pixels might catch some of what this vocabulary misses. It would also require more examples and speak back less clearly. Here, blindness is accepted in exchange for legibility: the machine's account of taste remains available for argument. A hybrid can come later. First it is useful to know how far an authored vocabulary can carry us, and where it breaks.

4 Learning within a changing practice

4.1 A progressively complex preference model

For rated recipes $\{(c_i, y_i)\}_{i=1}^{n}$, the scoring model learns the plain relation

$$f : \phi(c_i) \mapsto y_i, \qquad y_i \in \{-1.5, 0.5, 2.5\}$$

between a machine-readable recipe and the judgment it received. With fewer than eight usable memories, the implementation falls back to ridge regression. After that it uses histogram gradient-boosted regression trees from scikit-learn (Pedregosa et al. 2011). The trees are allowed to grow more articulate as the archive grows: depth one below 800 ratings, depth two through 1,999, and depth three thereafter. As such, the ontology becomes richer as the data set grows.

Each tier changes the kind of sentence the model can form. At depth one it can learn that felt tip fields tend to be loved. At depth two it can learn that they are loved in the monochrome palette. At depth three it can learn that they are loved there only with a particular background or placement. The thresholds are a studio heuristic, not an experimentally tuned law. The intention is simply to keep the machine from inventing elaborate reasons before it has accumulated enough experience to support them.

4.2 Exploration, inheritance, and diversity

The first 222 ratings—one for each feature dimension—come from fully random recipes. Until then, the model is kept out of the loop, so an uninformed account of taste cannot steer the search into premature convergence. Later candidates inherit from highly weighted ancestors through crossover and mutation: palettes change, traits appear or vanish, elements move, and gestures cross from one zone into another. Rendering details are stripped away so that each descendant must still prove itself on its own.

The engine generates four times as many candidates as the interface needs. It scores them all, but does not obediently return the highest numbers. Beginning with the strongest candidate, it adds the rest through a maximal-marginal-relevance-style compromise between predicted preference and difference (Goldstein and Carbonell 1998):

$$U(c) = 0.3\,\tilde{f}(c) + 0.7 \min_{s \in S} d(c, s),$$

where $\tilde{f}$ is the normalized model score and d measures difference from what has already been selected: categorical disagreement, trait-set distance, element counts, coverage, center, spread, and occupied zones. The 0.7 coefficient given to difference is deliberately large. Without it, a locally successful recipe quickly becomes a monoculture, and taste thrives on tension more than repetition.

Exploration is not noise around the real objective. It is part of the artistic position of the system. The machine must preserve the possibility that its artist will change their mind, that an apparent failure will open a lineage, or that the model's clearest account of current taste is precisely what needs to be resisted.

5 What the machine could recover

The following figures are generated directly from one local feedback database and recommender state. This database records one of multiple experiments in lineage-building and the deliberate artistic exploration of various facets of the vocabulary. In this experiment specifically, I committed to only rating tacitly, giving myself less than three seconds per image. While this invites misclassification, the experimental question was whether, given enough ratings, the machine and I would start to agree in tacit judgment.

5.1 Ratings and predictive fit

Between 25 March and 12 April 2026, the instrument accumulated 2,196 judgments: 344 *love*, 957 *meh*, and 895 *hate*. Love occupies only 15.7% of the archive. This is not a balanced benchmark, which is to be expected, since I do not think highly of most of my own work. It is best thought of as a record of a demanding, even gruelling encounter in which most outputs failed to become necessary. Correcting that imbalance would erase the actual pressure under which the generator evolved, and it would undermine aesthetic integrity.

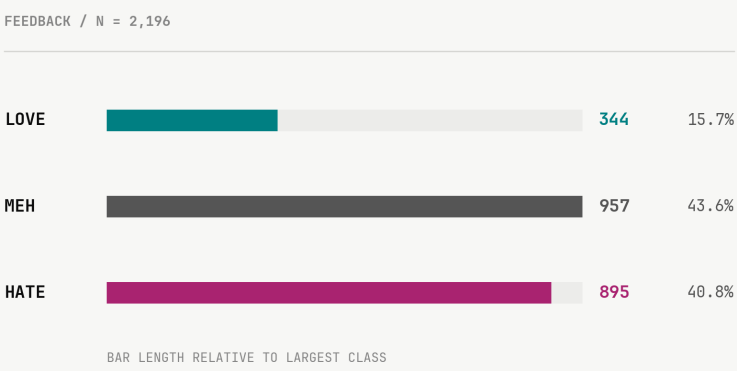

Figure 4. Rating distribution for the complete single-user dataset (*n* = 2,196).

At this size the model has reached depth three. Its training R^2 is 0.336. Shuffled three-fold cross-validation yields 0.225, 0.226, and 0.240: a mean of 0.230 with standard deviation 0.007. The number is neither miraculous nor frustrating. A little under a quarter of the held-out variation can be recovered from the decisions the ontology can name. A respectable if not prodigious batting average.

That claim needs to be examined carefully. Random folds mix early and late moments of an adaptive generator, and many recipes share ancestry. But the machine's earlier selves are not lost. Because each model is rebuilt from the archive, hiding the ratings that had not yet happened lets us reconstruct what it knew at any point in the encounter. Those earlier models recovered 22.4% of the variation in the next batches, nearly the same as the shuffled estimate of 23%.

The apparent stillness conceals changes, however. A model frozen after the first 222 ratings can recover 13.2% of the judgments that follow. Allowing it to keep learning raises that figure to 22.4%. Agreement does not steadily increase because the thing being learned is a moving target. The archive changes the generator, the generator changes what the artist encounters, and the artist may change in response. Learning does not carry the machine toward a final account of taste, it allows the machine to remain in a conversation. Concretely, my standards for *love* rise as I experience the quality of the average artwork increasing.

In-sample predicted scores are ordered by the observed labels: love has mean 0.631 (SD 0.488), meh 0.192 (SD 0.553), and hate −0.452 (SD 0.556). The ranges overlap substantially. This machine has not found a border between good and bad images. Instead, it has found a slope. That is enough to lean a batch toward more promising territory while still letting final artistic judgment have a meaningful impact.

label	*n*	mean score	SD	range
love	344	0.631	0.488	[−0.946, 1.805]
meh	957	0.192	0.553	[−1.398, 1.717]
hate	895	−0.452	0.556	[−1.600, 1.428]

Table 2. In-sample score distribution by observed rating. These values describe separation on the training set and are not held-out estimates.

5.2 A readable preference profile

When the fitted model is disturbed one feature at a time, one preference towers over the rest. The global count of the sand-creature trait has permutation importance 0.216, more than six times the next feature, global ink splatter at 0.035. Brush-stroke coverage (0.029), global bold slash (0.026), NightSlate palette (0.020), and charcoal backgrounds (0.017) follow. Center density, mean anchor size, and sand creatures placed in the center also recur near the top.

That dominance is easy to misread as a skew in taste. Sand creatures are not especially useful by themselves. Their scribbles add a depth to the texture that none of the other elements do, filling a void in the visual language. The feature's importance therefore exposes an ontological limitation: the vocabulary has only one way to say something the work repeatedly needs.

The raw archive tells a similar, if blurrier, story. NightSlate carries an average target weight of 0.446 across 299 appearances. Charcoal backgrounds average −0.287 across 526. The light "whisper" texture averages 0.165 across 376. These are associations within a tangled, self-altering process, not laws of cause and effect: features travel together and appear with unequal frequency.

Figure 5. Mean rating-target weight for located traits in the nine spatial zones. Support is shown in each cell. Counts and co-occurrence vary, so the map shows observed associations rather than isolated placement effects.

The center's repeated appearance is a signal that positioning is a meaningful trait. A bag of global traits would have discarded that signal. Yet the map should not be read as proof that the artist simply "likes the center." Placement, identity, composition, and ancestry remain entangled, and the presence of any sort of tension in the center of the image is all we can deduce from the figure. It tells a disjointed story if read on its own.

6 Where the ontology breaks

The machine can recover roughly 23% of held-out rating variance using 2,196 labels from one person. More importantly for practice, it can rank candidates well enough to change what appears next. This is possible because it does not begin with the impossible task of understanding an image. It operates purely on the compositional plane.

Composition is not innocent. Creating and distinguishing gestures, then dividing the page into nine zones, turns studio language into machine structure. The artist does not merely rate the outputs. They legislate, in advance, what may count as a reason. Feature engineering becomes aesthetic self-description.

This is why the system should not be mistaken for an automated connoisseur. It is a partial memory of components. Its readable preference profile can be accepted, resisted, or treated as a sinful provocation. Its diversity rule keeps open the possibility of surprise while still remaining firmly rooted in one practice. Its failures expose the edge of its language and point at subtle relations. When a seed produces an arrangement that matters for reasons no feature can explain, that is not statistical noise, it is an artwork created in the liminal space between machinery and ontology.

Again, figure 1d makes the problem concrete. Its materials and gestures were individually promising, but their relations produced too little tension, hierarchy, or movement. The ontology can name the parts without fully describing what passes between them. Taste may refuse quantification most stubbornly in those relations, and therefore precisely where the experiment becomes most interesting. Many of the most interesting results reside in the *hate* bucket. Together they compose the negative space of my personal taste.

7 Limits of this encounter

This archive belongs to one person, who is also the maker of the system. It says nothing about consensus, and it should not. The feature ontology belongs to one renderer. The three labels compress ambivalence, confidence, context, and changing mood. Their numeric weights are authored constants. The examples share ancestry, the sampling process adapts as it runs, and no amount of statistics can simulate the future of a practice, though we can recount its past. Feature importance, too, belongs to this model and this distribution.

Adding more features as the studio's practice grows is a direction that is artistically promising but methodologically uninteresting, since the task is purely mechanical. Still, this is the most likely direction in which the system might evolve in the future, since, in the end, the system is about artfulness.

A multi-user system would be worthwhile only if it refused the easy collapse into consensus. Each person should retain a separate, evolving model. The point is not to average private judgment into another generic aesthetic score, but to make difference visible over a shared field of possible drawings.

8 Reciprocal misunderstanding

As creation moves into algorithmic formality, the human contribution does not evaporate. It migrates. It appears in the world the artist permits the machine to inhabit, in the names given to that world's parts, in the refusals that prune it, and in the taste that guides it without ever becoming fully explicit.

This paper has given one such taste a partial memory. A procedural drawing is represented through 222 authored decisions; three quick judgments train a bounded preference surface; evolutionary variation and deliberate diversity turn that surface back into new propositions. The model learns enough to steer without learning enough to settle the matter. The artist remains as ontologist and curator.

That incompleteness is not a defect to be engineered away. It keeps visible the distance between the work and every vocabulary used to explain it. The machine remembers what the artist has noticed, named, and chosen so far. The artist meets that memory in the next batch and may confirm it, contradict it, or become someone it no longer describes. Somewhere in that reciprocal misunderstanding, the new creative act struggles to be born.

For now, it remains a co-creative monster.

References

Frans, Kevin, L. B. Soros, and Olaf Witkowski (2022). "CLIPDraw: Exploring Text-to-Drawing Synthesis through Language–Image Encoders". In: *Advances in Neural Information Processing Systems*. Vol. 35. arXiv: 2106.14843.

Fürnkranz, Johannes and Eyke Hüllermeier, eds. (2010). *Preference Learning*. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-14125-6.

Goldstein, Jade and Jaime Carbonell (1998). "Summarization: (1) Using MMR for Diversity-Based Reranking and (2) Evaluating Summaries". In: *TIPSTER Text Program Phase III: Proceedings of a Workshop*, pp. 181–195. DOI: 10.3115/1119089.1119120.

Gramsci, Antonio (1971). *Selections from the Prison Notebooks of Antonio Gramsci*. Ed. and trans. by Quintin Hoare and Geoffrey Nowell Smith. London: Lawrence and Wishart.

Lv, Pei, Meng Wang, Yongbo Xu, Ze Peng, Junyi Sun, Shimei Su, Bing Zhou, and Mingliang Xu (2018). "USAR: An Interactive User-specific Aesthetic Ranking Framework for Images". In: DOI: 10.48550/arXiv.1805.01091. arXiv: 1805.01091 [cs.CV].

McCormack, Jon, Oliver Bown, Alan Dorin, Jonathan McCabe, Gordon Monro, and Mitchell Whitelaw (2014). "Ten Questions Concerning Generative Computer Art". In: *Leonardo* 47.2, pp. 135–141. DOI: 10.1162/LEON_a_00533.

McCormack, Jon and Andy Lomas (2020). "Deep Learning of Individual Aesthetics". In: *Neural Computing and Applications* 33.1, pp. 3–17. DOI: 10.1007/s00521-020-05376-7.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12.85, pp. 2825–2830.

Takagi, Hideyuki (2001). "Interactive Evolutionary Computation: Fusion of the Capabilities of EC Optimization and Human Evaluation". In: *Proceedings of the IEEE* 89.9, pp. 1275–1296. DOI: 10.1109/5.949485.

Talebi, Hossein and Peyman Milanfar (2018). "NIMA: Neural Image Assessment". In: *IEEE Transactions on Image Processing* 27.8, pp. 3998–4011. DOI: 10.1109/TIP.2018.2831899.

Žižek, Slavoj (2010). "A Permanent Economic Emergency". In: *New Left Review* 64, pp. 85–95. URL: https://newleftreview.org/issues/ii64/articles/slavoj-zizek-a-permanent-economic-emergency.