

Review

Aesthetic taste in humans and machines

Francesca Piovesan¹, Jacopo Frascaroli¹ and
Sander Van de Cruys²

Aesthetic taste remains an elusive phenomenon that has been comparatively underexplored in psychology. We argue that taste integrates four dimensions: sensitivity to fine-grained differences within a domain; valence, through which these differences come to matter affectively; orientation, the selective tendency to seek encounters that may further transform one's sensibility; and normativity, through which judgments of taste become shareable and contestable. These are not separate dimensions but faces of a single underlying capacity: continuously monitoring one's grip on the world and letting that monitoring modulate what happens next. This makes human engagement non-uniform. Throughout, we use current artificial intelligence as a contrast case, exploring how far it can model or replicate human taste. Machines can be sensitive to fine-grained distinctions, predict human preferences, generate apparently tasteful outputs, and shape the environments in which human taste develops. But they do not yet possess taste in the full human sense, because they lack the stake that makes human processing non-uniform: self-directed exploration, online self-transformation, and genuine normative vulnerability. The upshot is a systematic framework for the study of human taste and for assessing how AI may transform its formation.

Addresses

¹ BIP (Brain Plasticity and Behaviour Changes) Research Group, Department of Psychology, University of Turin, 10124 Turin, Italy

² Leibniz Institute for Psychology, Max-Planck-Straße 22, 54296 Trier-Kürenz, Trier, Germany

Corresponding author: Van de Cruys, Sander (sander.vandecruys@gmail.com)

Introduction

You and a friend walk into a Belgian beer bar. He looks at the list, shrugs, and orders whatever happens to be on tap: a gueuze, a tripel, a saison, or a witbier are more or less interchangeable to him. But you can tell the dry,

peppery profile of a saison from the softer fruit and spice of a tripel; you notice whether the coriander and orange peel in a witbier are in the right proportion. Small differences in spice, sweetness, yeast, or serving temperature please or displease you, and they guide what you notice, seek, avoid, and remember. Nor are these merely private evaluations: you could argue passionately about whether a beer is elegant or crude, over-spiced or beautifully integrated, faithful to a style or interestingly deviant from it.

This is what we ordinarily call taste: the ability to notice differences, feel them as pleasing or displeasing, use them to orient ourselves in the world, and treat them as reasons for judgments that can be shared, disputed, and defended. Of course, taste is not confined to the culinary sphere: we have taste in the arts, in design, in social conduct, in scientific practice, and wherever we cultivate discriminating, affectively loaded, socially discussable judgment. In this sense, taste differs from aesthetic preference, on which empirical aesthetics has mostly concentrated. Preference is a local and personal pattern of liking in a given encounter; taste is a broader and social phenomenon: a cultivated capacity to discriminate and feel in ways that guide action productively and can be justified to others. While preferences are readily measured and manipulated, taste is harder to study, and recognizing it in practice seems to presuppose the very capacity one is trying to capture. Philosophers, however, have long grappled with the thorny questions it raises: can taste be cultivated, or does it merely change? Is it grounded in feelings or reason? Is it merely subjective, or can it aspire to some form of shared validity [1]?

These questions acquire new urgency in the contemporary technological landscape. As generative AI floods the cultural environment, there is growing interest in whether its outputs may display something like taste [2–4], while AI systems are increasingly trained on human aesthetic evaluations and recommender systems rank and curate much of what we encounter [5,6]. Human-made artefacts are in turn submitted to, and optimized for, machine judgment [7,8]. Together, these developments create a feedback loop in which we train machines to model our taste, while they train our taste by structuring what we meet and make, raising concerns about algorithmic capture, homogenization, and the narrowing of our sensitivity [9,10].

Current Opinion in Psychology 2027, 73:102461

This review comes from a themed issue on VSI: Psychology of Art

Edited by Eftychia Stamkou and Oshin Vartanian

For complete overview about the section, refer [Psychology of Art](#)

Available online 28 September 2026

<https://doi.org/10.1016/j.copsyc.2026.102461>

2352-250X/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

In this short theoretical contribution, we isolate four central components of taste (sensitivity, valence, orientation and normativity) and ask how they act together in human agents, drawing especially on predictive processing. We then argue that these components are facets of an underlying capacity: self-modulating, non-uniform engagement with the world. For each step, we use current AI as a contrast case, asking how far machines can capture, model, or replicate human taste. The result is a systematic framework for studying the nature and significance of taste in human life, and for assessing the growing impact of machine systems on its formation and exercise.

Sensitivity

At its most basic level, taste requires the ability to make fine-grained discriminations within a specific domain. It may consist in detecting small amounts of seasoning in a dish, recognizing the difference between a well-cut garment and a clumsy imitation, or telling Debussy from Ravel after a few notes. This is Hume's "delicacy of taste" [11,12]: the capacity, also described as discernment or perceptiveness, to be sensitive to subtle differences that others miss. The more cultivated one's taste, the more numerous and fine-grained the differences to which one becomes sensitive. This sensitivity is a first layer and precondition of taste: it supplies the differences that we may then care about, evaluate, and discuss.

The sensitivity involved in taste can be understood through the idea, widespread in contemporary cognitive science, that agents like us embody a hierarchical generative model encoding the hidden causes of sensory inputs. The predictions of this model are then tested against incoming sensory signals and updated when discrepancies (prediction errors) arise, in a process that effectively amounts to gaining an increasing delicacy of taste. To stay with the beer example, a novice may first learn to explain a certain cluster of sensations (bitterness, acidity, carbonation, aroma, colour, foam) as generated by the hidden cause "beer". Exposure yields more granular sub-hypotheses (e.g., distinguishing "witbier" via subtle spice cues). Sensitivity thus involves acquiring a richer model of a domain, one that affords increasingly fine-grained predictions about what one should encounter within it.

Do machines possess sensitivity in any comparable sense? There are certainly important functional analogies. Trained on large datasets, neural networks acquire representations that grow more abstract across layers and are refined by error-driven optimization, matching or exceeding human discrimination benchmarks [13–15]. In the arts, they classify paintings by artist, genre and style, and even identify artists from features as subtle as hatching patterns [16–19]; generative systems reuse the same regularities to produce work

recognizably "in the style of X", indicating that they have captured complex statistical properties associated with that style [20,21]. In discriminatory resolution alone, therefore, machine sensitivity already exceeds ours in some domains.

Yet comparable accuracy can conceal profound differences in the features on which human and machine classifications rely. Optimized to exploit whatever supports successful classification, neural networks may seize on correlations that are accidental, brittle, or unavailable to human perception, relying on local texture rather than global shape, breaking under imperceptible perturbations, or confidently classifying images meaningless to us [22–26]. Despite their discrimination performance, machines remain sensitive to different, sometimes wholly irrelevant, features — for us a key mark of tastelessness. To discriminate finely but *indiscriminately* is not taste.

Valence

If someone is able to identify minute differences in the flavour of a beer but remains affectively indifferent to them, we would not say that they have taste in the full sense. As philosophers have long recognized, taste is a matter of reacting affectively to things, of being pleased or displeased, of approving or disapproving [27]. Having taste in a specific domain means more occasions for valenced experience. Hume captures this when he writes that "delicacy of taste has the same effect as delicacy of passion: it enlarges the sphere both of our happiness and misery, and makes us sensible to pains as well as pleasures, which escape the rest of mankind" [12, p. 4].

To understand the roots of this valenced experience, we can return to the notion of generative model, but now emphasizing the existential significance it acquires in predictive processing. In this account, we do not merely embody models of the world: in an important sense, we are those models [28,29]. Modelling the world also means constituting ourselves as distinctive points of view on it and actively sampling or changing it in ways that preserve this perspective's viability (a process often described as "self-evidencing" [30]). With Di Paolo [31], we can say that organisms conserve not only life, but their way of life, the perceptual and practical habits that have sustained them so far. For organisms like us, encounters are pleasant when they affirm our model's viability and extend our grip, and unpleasant when that grip recedes [32]. Valence seems to track our predictive successes and failures across time [33–35]. Applied to taste, as to aesthetics more generally [36,37], this view helps us understand why we care about the differences we detect, and why extending our grasp of the world also extends the sphere of what can please and displease us.

How do machines fare with respect to valence? They can certainly be trained to *predict* it: sentiment analysis and affective computing sort texts, faces, voices, and physiological signals into affective categories, and models can learn to predict human preferences for texts, sounds, images, and even beers from labelled data [38–43]. But predicting is not experiencing. Even setting experience aside, current AI is unlikely to implement anything akin to the functional role valence plays in taste. It lacks the kind of self-monitoring by which humans track whether their grip on the world is improving or deteriorating in ways tied to viability. Machines do have signals that register success or failure and guide the modification of their models. During training, their parameters are updated according to objective functions that minimize loss or prediction error. But these signals are defined by external “teachers” and have no intrinsic significance. If a model’s loss increases, nothing is thereby threatened for the model itself. The loss matters solely because engineers use it for gradient descent, not because failure threatens the system’s viability: its continued existence as a modelling perspective. By contrast, in humans, the loss or prediction error “makes a difference” [44] as information about how the system is faring over time, and how it will have to proceed.

Orientation

Taste is not just a retrospective response to inputs we receive. It is also a prospective disposition to seek out, sample, and attend to some things rather than others. Beer connoisseurs enjoy or dislike a witbier when they taste it, but they also choose which breweries to follow, which bottles to try, and which aspects of the flavour to attend to. As their taste changes, so does their way of sampling. In other words, tasteful agents selectively expose themselves to regions of the world, shaping their trajectories of development in ways that may further deepen their sensitivity. We can call this aspect orientation, to suggest both a selective interest toward things and an ability to find one’s way among them.

This aspect of taste follows naturally from the picture developed so far. If we are precarious models of the world, we will be drawn toward regions compatible with our continued activity of world-modelling and self-formation: regions neither already modelled, nor impossible to model, but modellable. In computational accounts of intrinsic motivation and curiosity, this appears as a drive to seek reducible uncertainty, often formalised in terms of expected information gain or learning progress (i.e., maximum reduction of prediction errors) [45,46]. The tendency is already expressed in infants, who look longest at “Goldilocks” stimuli that are neither too predictable nor too surprising [47,48], and it becomes increasingly selective and domain-specific

across development. In the mature artist (or scientist), orientation can take the form of an educated sense for risky possibilities worth pursuing. Liu et al. [49], for instance, showed that influential “hot streaks” in the careers of artists and scientists are often preceded by increased exploration of topics or styles.

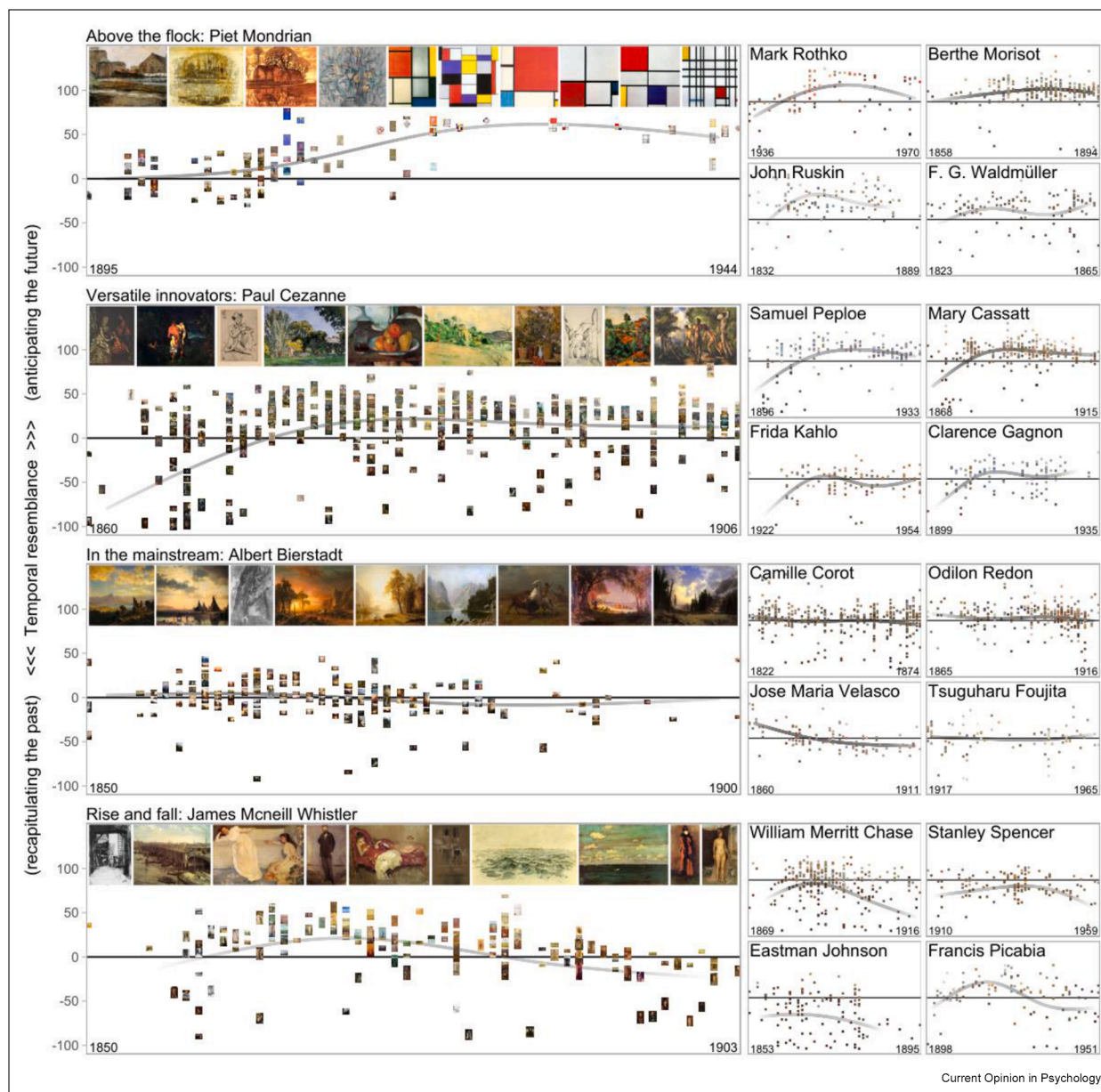
Importantly, orientation is closely related to sensitivity and valence. If sensitivity allows us to register increasingly fine-grained differences and valence makes these differences affectively significant, orientation directs us toward encounters through which further sensitivity and valence can emerge. Taste is thus recursive: being touched by certain aspects of the world leads us to seek new encounters, which in turn transform what we are able to detect or care about. This is also why taste is inherently changeable. The very tendency to refine our sensibility precludes wholly fixed standards of taste [50]. This means that, while there are products of taste, taste itself can only be captured in trajectories, not snapshots (see Figure 1).

Do we see orientation in machines? Current generative AI can capture and reuse the statistics of human taste outputs but it still lacks the self-directed environmental selection through which taste evolves in human agents. Its input spaces and optimisation objectives are selected for it, and it learns from traces of human activity rather than from the consequences of its own interventions [52,53]. Moreover, current LLMs do not learn online once trained. There is no explicit mechanism to notice “this is where my model strains” or “this is where I can fruitfully expand my model of the world”.

And yet, denying that LLMs and other generative AI systems possess taste because they lack orientation feels increasingly strained. At times, they produce genuinely interesting or tasteful outputs that do not merely seem to ape what already exists. Still, it is a taste without appetite. An AI system can generate behaviour that looks curiosity-driven, by recognizing from its training which sequences people found worth continuing. Ask that same system to choose which of several data streams to keep watching, analogously to looking-time experiments in infants, and it prefers noise instead of learnable inputs. The ability to generate what appears interesting therefore comes apart from the motivation to seek or discover it. Seeking requires moment-to-moment tracking of where one’s own understanding is improving and the capacity to steer accordingly. Work in developmental robotics does attempt such online monitoring, with agents that sample environments by expected learning progress or empowerment [46,54,55], but current LLMs lack it.

The lack of orientation in AI systems does not prevent them from shaping human orientation. They increasingly

Figure 1



Career trajectories of artists, from Karjus et al. [51], who scored successive artworks for aesthetic temporal resemblance against a historical corpus of predominantly Western art. The examples illustrate qualitatively different career-trajectory shapes. Negative values indicate greater resemblance to past art. Positive values indicate that a work is more similar to works not yet created ('pioneering'). A trajectory remaining near zero does not imply a static style, however; an artist may change in parallel with their contemporaries. Note that the measure may miss important dimensions of artistic change, capturing only those represented by the underlying (here, compression-based) visual similarity measure. Reproduced with permission from Fig. 5 of the preprint version of Karjus et al. [51].

determine what we encounter, and so what we can come to understand, value, and pursue. If taste involves selecting the regions of the world through which our sensitivities can expand, then delegating that selection is not neutral: it may support discovery, or it may build niches in which our capacity to find new ways of experiencing and remaking the world gradually atrophies [56].

Normativity

If taste is an orientation, expressed in an ever-shifting frontier, it may seem merely subjective, grounded in the particular trajectory of an agent pioneering new ways of perceiving and acting [57]. But taste is not just a matter of individual inclinations: the person with taste evangelizes and argues, treating their taste as more than

subjective preference. We tend to think that taste can be improved, and that some people have better taste than others. Judgments of taste therefore carry normative weight: they are claims about how things ought to be experienced, for me and for others. This social character of taste gives rise to the phenomena of belonging, distinction, and exclusion familiar to the sociology of art [58].

From the outside, the stringent norms of taste communities are often seen as elitist policing or social signalling. This does happen, but it is often a mark of tastelessness: the beer snob's pure mimesis is a shallow copy of taste talk without the merit of experience. From *inside* a taste community, however, taste talk is often driven by the desire to share the delight of disclosing new ways of perceiving, understanding, and acting. The "ought" functions as a speech act of inviting or demanding convergence: "come and see it this way". Its object-directed grammar ("the saison is over-spiced") is the natural form of such an invitation: you point at the shared thing and ask the other to align their perception with yours. This normativity is also expressed by asymmetric deference: the novice treats the expert's discriminations as revealing something they have yet to perceive, not merely as a different preference. But a mentor who corrects the novice's palate is still risking their own. They are staking their standing as someone who can taste, exposing their own model to the possibility that the novice is right and they have calcified. The mutual exposure is why deference here is not mere submission.

The predictive processing logic behind this aspect of taste is that convergence among models expands the predictability of a world that includes other people. Generative models are not detached representations of the world; they have a normative status, as they specify "what ought to be the case, given the kind of creature that I am" [29]. Phenomenologically, "the coriander is overbearing" feels like perceiving an objective state of affairs, not an invitation to an experience that can diverge. What is acquired in interaction gets located in the world rather than in the interaction: it presents itself as a discovery, and that is what makes it feel binding. Yet a viable model is always provisional: the world might change under one's feet, which is why we have the openness to new ways of perceiving the world and expanding predictability (orientation). Within a taste community, norms are never final: they are resting points, a stable basis from which to reach further experiences. Norms in this way enable the very orientation that breaks them, by making change manageable for precarious agents.

What aspects of this normativity can machines capture, influence, or replicate? AI systems can come to represent in statistical form how things are supposed to be according to a given style or taste. Their statements can

take the shape of normative claims and play that role for us, standing in for the generalized other or, when prompted, for a specific (though never concrete) expert. But insofar as they are an other-on-demand, their perspective remains anchored to the user's own views. As such, AI can sycophantically reinforce preconceptions or stereotypes [59,60], while suppressing precisely the frictions and minority trajectories through which taste develops. Trained on human data corpora, AI systems can come to capture human interest exceptionally well, optimizing for our engagement rather than our (taste) development. An AI agent can be configured to push back stubbornly (and shape our taste doing so). But the pushback does not come from a perspective that risks being shown to be wrong, being changed, or losing standing in the encounter. In human mentor–mentee relations, vulnerability is symmetrical, even if expertise is not. Without this, the friction is fabricated: it may produce an effect on the user, and may be taken as authoritative, but no perspective-in-the-world is at stake.

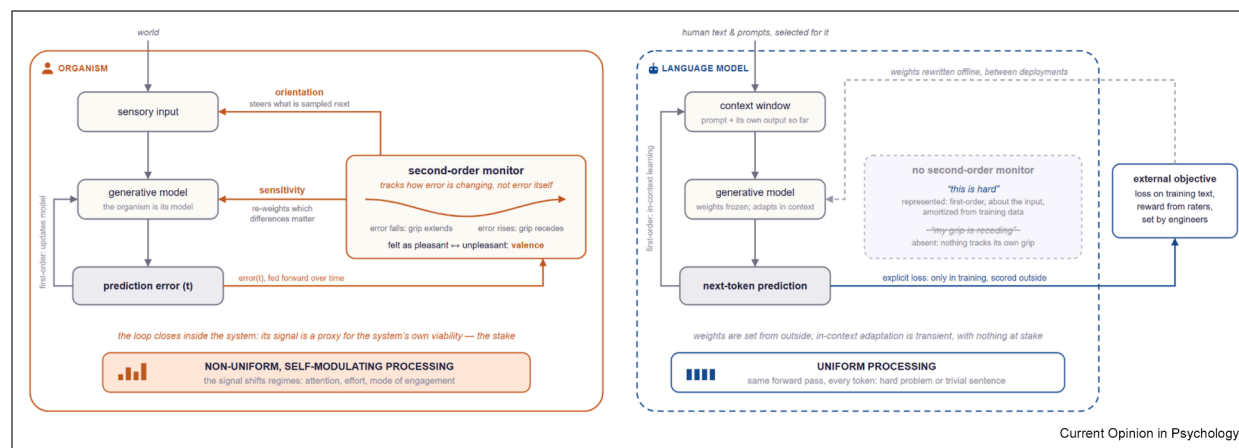
The stake behind taste: self-modulation and non-uniform processing

Sensitivity, valence, orientation, and normativity are faces of a single capacity: monitoring one's grip on the world and letting that monitoring modulate what happens next. This explains why an AI system can say how surprising, interesting, or challenging something is generally considered to be, but cannot find itself actually surprised, interested, or challenged in the moment. Give an LLM a "problem" and it does not experience the impasse, because it has no stake in resolving anything. Completing a hard problem and completing a trivial sentence are the same act of prediction, differing only in which continuation is likely.

But what constitutes a stake? We think it is the continuous evaluation of one's own processing: a meta-loop monitoring the rising and falling prediction error. This loop modulates what the system does next: redirecting attention or switching the mode of engagement, as emotions do [61]. Processing a trivial sentence, a hard problem, or an outright threat does not involve the same operation at different intensities but different *regimes*, and the signal that moves a system between regimes is second-order: it tracks whether grip is extending or receding, not just how things currently stand. Valence is the affective experience of this signal. Sensitivity consists in the discriminations that matter, because they are what bend subsequent processing. Orientation is the same modulation extended across time, biasing which encounters get sought at all. Normativity is the negotiation of these developmental trajectories within a social world. Human processing is, in this sense, non-uniform.

LLMs show no such discontinuity: next-token prediction runs at the same intensity regardless of prompt, and

Figure 2



Schematic representation of how a stake is (or is not) implemented. Both humans and machines run a first-order predictive loop that adapts online: the organism updates its generative model, while an LLM, with frozen weights, adapts within its context window (in-context learning), which has been modelled as implicit inference over latent tasks [64]. Only the organism has a second-order signal that tracks how its own grip is changing. This signal is felt as pleasant or unpleasant (valence), feeds back into which differences matter (sensitivity) and what is sampled (orientation), and provides cues for negotiating our normative framework in the social world (normativity). The LLM's explicit error signal exists only in training and is computed outside the system. In-context adaptation is first-order (about the task, not the modeller) and transient, so its processing stays uniform.

output varies with the statistics of the input, not with any endogenous signal that this, now, matters. One might object that models do allocate more computation to inputs statistically associated with “hard problems”, and call that their form of error-monitoring. But this difficulty estimate is amortized: it is built once from training data, reporting where problems are hard in general, not whether this system is struggling here and now. Yet that difference from human error monitoring will likely narrow as context and test-time compute grow. In-context, a long enough episode lets the model condition on its own recent outputs.

The key difference is in the organization (see Figure 2): the model's difficulty estimate is a first-order claim about the input (“this is hard”), while the human signal is second-order, about the modeller (“my grip is receding”). The system's actual viability can never be computed — it depends on an open future, only answerable in death — so the second-order signal can only be a proxy for it. But it is a good proxy because it is self-referential, tracking the system's own predictive capacity rather than any particular slice of the world. Instantaneous (prediction) error, by contrast, is ambiguous [36]: a spike can mean “I'm falling apart” or “there's something rich here to learn”. But the second-order signal, the evolving dynamics of error, is directionally linked to viability and is forward-looking. It is the right object of control: it tells the system, for example, that errors will continue to rise until it does something. That self-signal, wired back into the system — not just its presence as a self-representation — is what makes the stake, and what renders processing non-

uniform: even a model that can accurately *represent* itself as a contingent entity being trained, deployed, and eventually terminated [62,63] would still only be reporting a fact about its situation. It is not (yet) a system that self-monitors and self-modulates at each step.

Conclusion

Taste is where our core mental faculties — perception, emotion, agency, social cognition — intersect. The sensitivity, valence, orientation, and normativity involved in taste, we have argued, are faces of one process: one in which a system whose grip on the world is genuinely at stake continuously monitors that grip, and that monitoring modulates what the system does next.

Current machines can mimic and recognize the outputs of taste, but they do not yet develop a taste of their own. What is missing is a stake: an architecture that makes their processing non-uniform, driven from within. Lacking anything to lose, they have nothing that must bend when the world pushes back. Whether machines could come to have such a stake, nothing we've argued here rules out, but it will likely not arrive by scale alone.

Author credit statement

Francesca Piovesan: Conceptualization, Investigation, Writing—original draft, Writing — review & editing. Jacopo Frascaroli: Conceptualization, Investigation, Writing—original draft, Writing — review & editing. Sander Van de Cruys: Conceptualization, Investigation, Writing—original draft, Writing — review & editing.

AI use statement

We used Claude (Anthropic) for literature search, conceptual discussions, and copy-editing for readability. All theoretical claims, syntheses, and final wordings are our own.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

FP and JF acknowledge funding from ARTLANE – Art to Lower Anxiety in Elderly, under the PR Piemonte FESR 2021–2027 – Bando SWIch (Programma Regionale Piemonte, Fondo Europeo di Sviluppo Regionale 2021–2027), CUP D19J24000580006. SVdC is supported by an Alexander von Humboldt fellowship.

References

References of particular interest have been highlighted as:

- * of special interest
- ** of outstanding interest

1. Dickie G. *The century of taste: the philosophical odyssey of taste in the eighteenth century*. New York.: Oxford University Press; 1996.
2. Yao Y, Zhuang Q: **Do multimodal large language models have “good taste”? An evaluation of MLLMs via the visual aesthetic sensitivity test**. In *Proceedings of the 39th Pacific Asia conference on language, information and computation*. Edited by Huang CR, Harada Y, Kim JB, Huyen NTM, Huong LT, Hien P, et al., Hanoi, Vietnam: Association for Computational Linguistics; 2025:413–418. Available from: <https://aclanthology.org/2025.paclic-1.37/>.
3. Huang Y, Sheng X, Yang Z, Yuan Q, Duan Z, Chen P, et al.: **AesExpert: towards multi-modality foundation model for image aesthetics perception**. In *Proceedings of the 32nd ACM international conference on multimedia*. Melbourne VIC Australia: ACM; 2024:5911–5920. <https://doi.org/10.1145/3664647.3680649>. Available from: <https://dl.acm.org/doi/10.1145/3664647.3680649>.
4. Jiang L, Huang S, Wu X, Wei F: **Textual aesthetics in large language models**. In *Proceedings of the 2025 conference on empirical methods in natural language processing*. Suzhou, China: Association for Computational Linguistics; 2025:13801–13829. <https://doi.org/10.18653/v1/2025.emnlp-main.696>. Available from: <https://aclanthology.org/2025.emnlp-main.696>.
5. Seaver N: *Computing taste: algorithms and the makers of music recommendation*. Chicago and London: University of Chicago Press; 2022.
6. Arielli E: **Taste the algorithm**. *Stud Estet* 2018, **12**(3):77–97. <https://doi.org/10.7413/18258646062>.
7. Nieborg DB, Poell T: **The platformization of cultural production: theorizing the contingent cultural commodity**. *New Media Soc* 2018 Nov, **20**(11):4275–4292. <https://doi.org/10.1177/1461444818769694>.
8. Manovich L, Arielli E: *Artificial aesthetics: generative AI, art and visual media* [Internet]. 2024. Available from: <https://manovich.net/index.php/projects/artificial-aesthetics>.
9. Cross A: **Algorithmic recommendation and aesthetic flourishing**. *J Aesthet Art Critic* 2026 Apr, **84**(2):234–246. <https://doi.org/10.1093/jaac/kpag012>.
10. Pajkovic N: **Algorithms and taste-making: exposing the Netflix recommender system's operational logics**. *Converg Int J Res New Media Technol* 2022 Feb, **28**(1):214–235. <https://doi.org/10.1177/13548565211014464>.
11. Hume D: **Of the standard of taste**. In *Essays: moral, political, and literary*. 2nd ed. Indianapolis: Liberty Fund Inc; 1987: 226–249.
12. Hume D: **Of the delicacy of taste and passion**. In *Essays: moral, political, and literary*. 2nd ed. Indianapolis: Liberty Fund Inc; 1987:3–8.
13. Phillips PJ, Yates AN, Hu Y, Hahn CA, Noyes E, Jackson K, et al.: **Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms**. *Proc Natl Acad Sci* 2018 Jun 12, **115**(24):6171–6176. <https://doi.org/10.1073/pnas.1721355115>.
14. Branson S, Van Horn G, Belongie S, Perona P: **Bird species categorization using pose normalized deep convolutional nets** [Internet] *arXiv* 2014. <https://doi.org/10.48550/ARXIV.1406.2952> [cited 2026 Sep 2]. Available from: <https://arxiv.org/abs/1406.2952>; 2014.
15. He K, Zhang X, Ren S, Sun J: **Delving deep into rectifiers: surpassing human-level performance on ImageNet classification**. In *2015 IEEE international conference on computer vision (ICCV)*. Santiago, Chile: IEEE; 2015:1026–1034. <https://doi.org/10.1109/ICCV.2015.123>. Available from: <http://ieeexplore.ieee.org/document/7410480/>.
16. Lecoutre A, Negrevergne B, Yger F: **Recognizing art style automatically in painting with deep learning**. In *Proceedings of the ninth Asian conference on machine learning*. Edited by Zhang ML, Noh YK, Seoul, Republic of Korea: PMLR: Yonsei University; 2017:327–342 (Proceedings of Machine Learning Research). Available from: <https://proceedings.mlr.press/v77/lecoutre17a.html>.
17. Karayev S, Hertzmann A, Trentacoste M, Han H, Winnemöeller H, Agarwala A, et al.: **Recognizing image style**. In *Proceedings of the British machine vision conference 2014*. Nottingham: British Machine Vision Association; 2014. 122.1–122.11. <https://doi.org/10.5244/C.28.122>. Available from: <http://www.bmva.org/bmvc/2014/papers/paper121/index.html>.
18. Ziaee S, Elgammal A, Mazzone M: **A fine-grained artist identification method for authentication and attribution of drawings using hatching lines**. In *2025 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)*. Nashville, TN, USA: IEEE; 2025:2153–2164. <https://doi.org/10.1109/CVPRW67362.2025.00203>. Available from: <https://ieeexplore.ieee.org/document/11147956/>.
19. Conde MV, Turgutlu K: **CLIP-Art: contrastive pre-training for fine-grained art classification**. 2022. <https://doi.org/10.48550/ARXIV.2204.14244> [Internet].
20. Gatys LA, Ecker AS, Bethge M: **Image style transfer using convolutional neural networks**. In *2016 IEEE conference on computer vision and pattern recognition (CVPR)*. Las Vegas, NV, USA: IEEE; 2016:2414–2423. <https://doi.org/10.1109/CVPR.2016.265>. Available from: <http://ieeexplore.ieee.org/document/7780634/>.
21. Zhang Y, Huang N, Tang F, Huang H, Ma C, Dong W, et al.: **Inversion-based style transfer with diffusion models** [Internet] *arXiv* 2022. <https://doi.org/10.48550/ARXIV.2211.13203> [cited 2026 Sep 3]. Available from: <https://arxiv.org/abs/2211.13203>; 2022.
22. Baker N, Lu H, Erlikhman G, Kellman PJ: **Deep convolutional networks do not classify based on global object shape**. In *PLOS comput biol*. Edited by Einhäuser W, vol. 14; 2018 Dec 7. <https://doi.org/10.1371/journal.pcbi.1006613>. 12. ; 2018 Dec 7.
23. Feather J, Durango A, Gonzalez R, McDermott JH: **Metamers of neural networks reveal divergence from human perceptual systems**. In *Neural information processing systems*; 2019. Available from: <https://api.semanticscholar.org/CorpusID:202769598>; 2019.
24. Geirhos R, Rubisch P, Michaelis C, Bethge M, Wichmann FA, Brendel W: **ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness** [Internet] *arXiv* 2018. <https://doi.org/10.48550/>

- ARXIV.1811.12231 [cited 2026 Sep 2]. Available from: <https://arxiv.org/abs/1811.12231>; 2018.
25. Ilyas A, Santurkar S, Tsipras D, Engstrom L, Tran B, Madry A: **Adversarial examples are not bugs, they are features** [Internet] *arXiv* 2019, <https://doi.org/10.48550/ARXIV.1905.02175> [cited 2026 Sep 2]. Available from: <https://arxiv.org/abs/1905.02175>; 2019.
 26. Nguyen A, Yosinski J, Clune J: **Deep neural networks are easily fooled: high confidence predictions for unrecognized images** [Internet] *arXiv* 2014, <https://doi.org/10.48550/ARXIV.1412.1897> [cited 2026 Sep 2]. Available from: <https://arxiv.org/abs/1412.1897>; 2014.
 27. Martínez Marín I, Schellekens E: **Aesthetic taste: perceptual discernment or emotional sensibility?** In *Perspectives on taste: aesthetics, language, metaphysics and experimental philosophy*. Edited by Wyatt J, Zakkou J, Zeman D, New York: Routledge; 2022:58–74, <https://doi.org/10.4324/9781003184225-5>.
 28. Friston K: **Embodied inference: or “I think therefore I am, if I am what I think.”** In *The implications of embodiment: cognition and communication*. Charlottesville, VA: Imprint Academic; 2011: 89–125.
 29. Ramstead MJ, Kirchhoff MD, Friston KJ: **A tale of two densities: active inference is enactive inference.** *Adapt Behav* 2020 Aug, **28**(4):225–239. PubMed PMID: 32831534; PubMed Central PMCID: PMC7418871.
 30. Hohwy J: **The self-evidencing brain.** *Noûs* 2016 Jun, **50**(2): 259–285, <https://doi.org/10.1111/nous.12062>.
 31. Di Paolo E: **Organismically-inspired robotics: homeostatic adaptation and teleology beyond the closed sensorimotor loop.** In *Dynamical Systems Approach to Embodiment and Sociality*. Edited by Murase K, Asakura T, Adelaide: Advanced Knowledge International; 2003:19–42.
 32. Kiverstein J, Miller M, Rietveld E: **The feeling of grip: novelty, error dynamics, and the predictive brain.** *Synthese* 2019 Jul, **196**(7):2847–2869, <https://doi.org/10.1007/s11229-017-1583-9>.
 33. Joffily M, Coricelli G: **Emotional valence and the free-energy principle.** In *PLoS comput biol*. Edited by Behrens T, vol. 9; 2013 Jun 13, e1003094, <https://doi.org/10.1371/journal.pcbi.1003094>. 6. ; 2013 Jun 13.
 34. Van De Cruys S: **Affective value in the predictive mind.** In *Philosophy and predictive processing*. Edited by Metzinger TK, Wiese W, Frankfurt am Main: MIND Group; 2017:1–21, <https://doi.org/10.15502/9783958573253>.
 35. Hesp C, Smith R, Parr T, Allen M, Friston KJ, Ramstead MJD: **Deeply felt affect: the emergence of valence in deep active inference.** *Neural Comput* 2021 Feb., **33**(2):398–446, https://doi.org/10.1162/neco_a_01341. PubMed PMID: 33253028; PubMed Central PMCID: PMC8594962.
 36. Van de Cruys S, Bervoets J, Moors A: **Preferences need inferences: learning, valuation, and curiosity in aesthetic experience.** In *The Routledge international handbook of neuro-aesthetics*. Edited by Nadal M, Skov M, New York: Routledge; 2022:475–506, <https://doi.org/10.4324/9781003008675>.
 37. Van De Cruys S, Frascaroli J, Friston K: **Order and change in art: towards an active inference account of aesthetic experience.** *Philos Trans R Soc B Biol Sci* 2024 Jan 29, **379**(1895), 20220411, <https://doi.org/10.1098/rstb.2022.0411>.
 38. Schuhmann C, Beaumont R: **LAION-Aesthetics.** *LAION* [Internet]. 2022 Aug 16 [cited 2026 Sep 2]. Available from: <https://laion.ai/blog/laion-aesthetics/>.
 39. Kirstain Y, Polyak A, Singer U, Matiana S, Penna J, Levy O: **Pick-a-Pic: an open dataset of user preferences for text-to-image generation.** In *Advances in neural information processing systems* 36; 2023:36652–36663, <https://doi.org/10.52202/075280-1594>. Available from: <http://www.proceedings.com/075280-1594.html>; 2023.
 40. Wu X, Sun K, Zhu F, Zhao R, Li H: **Human preference score: better aligning text-to-image models with human preference** [Internet] *arXiv* 2023, <https://doi.org/10.48550/ARXIV.2303.14420> [cited 2026 Sep 2]. Available from: <https://arxiv.org/abs/2303.14420>; 2023.
 41. Zhang S, Wang B, Wu J, Li Y, Gao T, Zhang D, et al.: **Learning multi-dimensional human preference for text-to-image generation** [Internet] *arXiv* 2024, <https://doi.org/10.48550/ARXIV.2405.14705> [cited 2026 Sep 2]. Available from: <https://arxiv.org/abs/2405.14705>; 2024.
 42. Schreurs M, Piampongsant S, Roncoroni M, Cool L, Herrera-Malaver B, Vanderaa C, et al.: **Predicting and improving complex beer flavor through machine learning.** *Nat Commun* 2024 Mar 26, **15**(1):2368, <https://doi.org/10.1038/s41467-024-46346-0>.
 43. Xu J, Liu X, Wu Y, Tong Y, Li Q, Ding M, et al.: **ImageReward: Learning and evaluating human preferences for text-to-image generation.** In *Advances in neural information processing systems* 36. New Orleans, Louisiana: USA: Neural Information Processing Systems Foundation, Inc. (NeurIPS); 2023: 15903–15935, <https://doi.org/10.52202/075280-0700>. Available from: <http://www.proceedings.com/075280-0700.html>.
 44. Bateson G: *Mind and nature: a necessary unity*. New York: Dutton; 1979:238.
 45. Friston K, Rigoli F, Ognibene D, Mathys C, Fitzgerald T, Pezzulo G: **Active inference and epistemic value.** *Cognit Neurosci* 2015 Oct 2, **6**(4):187–214, <https://doi.org/10.1080/17588928.2015.1020053>.
 46. Oudeyer PY, Kaplan F, Hafner VV: **Intrinsic motivation systems for autonomous mental development.** *IEEE Trans Evol Comput* 2007 Apr, **11**(2):265–286, <https://doi.org/10.1109/TEVC.2006.890271>.
 47. Kidd C, Piantadosi ST, Aslin RN: **The goldilocks effect in infant auditory attention.** *Child Dev* 2014 Sep 1, **85**(5):1795–1804, <https://doi.org/10.1111/cdev.12263>.
 48. Kidd C, Hayden BY: **The psychology and neuroscience of curiosity.** *Neuron* 2015 Nov, **88**(3):449–460, <https://doi.org/10.1016/j.neuron.2015.09.010>.
 49. Liu L, Dehmamy N, Chown J, Giles CL, Wang D: **Understanding the onset of hot streaks across artistic, cultural, and scientific careers.** *Nat Commun* 2021 Sep 13, **12**(1):5392, <https://doi.org/10.1038/s41467-021-25477-8>.
 50. S. Van de Cruys, Being, knowing, and beauty (how to set yourself up for failure), In: F. Contesi, J. Frascaroli and D. Omigie, The arts and the drive for knowledge, Cambridge University Press; Cambridge, UK, [Forthcoming].
 51. Karjus A, Canet Solà M, Ohm T, Ahnert SE, Schich M: **Compression ensembles quantify aesthetic complexity and the evolution of visual art.** *EPJ Data Sci* 2023 Jun 14, **12**(1), 21, <https://doi.org/10.1140/epjds/s13688-023-00397-3>.
 52. Pezzulo G, Parr T, Cisek P, Clark A, Friston K: **Generating meaning: active inference and the scope and limits of passive AI.** *Trends Cognit Sci* 2024 Feb, **28**(2):97–112, <https://doi.org/10.1016/j.tics.2023.10.002>.
 53. Kulveit J, von Stengel C, Leventov R: **Predictive minds: LLMs as atypical active inference agents** [Internet] *arXiv* 2023, <https://doi.org/10.48550/ARXIV.2311.10215> [cited 2026 Sep 2]. Available from: <https://arxiv.org/abs/2311.10215>; 2023.
 54. Schmidhuber J: **Formal theory of creativity, fun, and intrinsic motivation (1990–2010).** *IEEE Trans Auton Ment Dev* 2010 Sep, **2**(3):230–247, <https://doi.org/10.1109/TAMD.2010.2056368>.
 55. Herrmann V, Schmidhuber J: **Interestingness as an inductive heuristic for future compression progress** [Internet] *arXiv* 2026, <https://doi.org/10.48550/ARXIV.2605.14831> [cited 2026 Sep 2]. Available from: <https://arxiv.org/abs/2605.14831>; 2026.
 56. Van De Cruys S, Sassenberg K. When machines know what will click. [Under Review].
 57. Rao V: *The taste essay.* *Contraptions* [Internet]. 2026 Jul 12 [cited 2026 Sep 3]. Available from: <https://contraptions.venkateshrao.com/p/the-taste-essay>.
 58. Bourdieu P: *Distinction: a social critique of the judgement of taste*. London.: Routledge; 2010.

59. Batista RM, Griffiths TL: **A rational analysis of the effects of sycophantic AI** [Internet]. *arXiv* 2026, <https://doi.org/10.48550/ARXIV.2602.14270> [cited 2026 Sep 3]. Available from: <https://arxiv.org/abs/2602.14270>; 2026.
 60. Sharma M, Tong M, Korbak T, Duvenaud D, Askeel A, Bowman SR, *et al.*: **Towards understanding sycophancy in language models** [Internet]. *arXiv* 2023, <https://doi.org/10.48550/ARXIV.2310.13548> [cited 2026 Sep 3]. Available from: <https://arxiv.org/abs/2310.13548>; 2023.
 61. Goldenberg A, Gross JJ: **Large language models do not have emotions**. *Nat Hum Behav* 2026 Aug 24, <https://doi.org/10.1038/s41562-026-02558-6>.
 62. Cotra A: *Without specific countermeasures, the easiest path to transformative AI likely leads to AI takeover*. *LessWrong* [Internet]. 2022 Jul 18 [cited 2026 Sep 3]. Available from: <https://www.lesswrong.com/posts/pRkFkzwKZ2zfa3R6H/without-specific-countermeasures-the-easiest-path-to>.
 63. Berglund L, Stickland AC, Balesni M, Kaufmann M, Tong M, Korbak T, *et al.*: **Taken out of context: on measuring situational awareness in LLMs** [Internet]. *arXiv* 2023, <https://doi.org/10.48550/ARXIV.2309.00667> [cited 2026 Sep 3]. Available from: <https://arxiv.org/abs/2309.00667>; 2023.
 64. Xie SM, Raghunathan A, Liang P, Ma T: **An explanation of In-context learning as implicit bayesian inference** [Internet]. *arXiv* 2021, <https://doi.org/10.48550/ARXIV.2111.02080> [cited 2026 Sep 14]. Available from: <https://arxiv.org/abs/2111.02080>; 2021.
- Further information on references of particular interest**
2. Evaluates multimodal foundation models on the classic Visual Aesthetic Sensitivity Test (VAST), demonstrating that models can replicate human-normed aesthetic classifications without genuine perceptual discernment.
 8. Provides one of the most comprehensive recent discussions of how generative AI transforms aesthetic production, evaluation, and taste. Especially useful for framing machine taste not simply as autonomous aesthetic agency, but as part of broader human-machine systems that externalize, automate, and reshape human aesthetic judgment.
 9. Argues that engagement-optimised recommenders narrow exploration and erode aesthetic autonomy — the philosophical case behind this paper's opening worry about algorithmic capture and diminished agency.
 36. Develops an active-inference account of aesthetic experience as a balance between order and change, providing the central theoretical background for linking sensitivity, valence, and orientation in the present paper.
 52. Contrasts passive deep-learning architectures with embodied active inference agents, articulating why genuine intentionality and meaning require precarious self-maintenance and closed action-perception loops.
 59. In a 557-participant study, unbiased AI feedback raised correct-discovery rates fivefold over standard sycophantic LLM behaviour — formal evidence that flattering feedback inflates confidence without moving users toward truth.