From Product to Producer: The Impact of Perceptual Evidence and Robot Embodiment on the Human Assessment of AI Creativity

NIKI PENNANEN, Department of Computer Science, Aalto University, Espoo, Finland and Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland
SIMO LINKOLA, Department of Computer Science, University of Helsinki, Helsinki, Finland
ANNA KANTOSALO, University of Helsinki, Helsinki, Finland
NICOLAS HIILLOS and TOMI MÄNNISTÖ, Department of Computer Science, University of Helsinki, Helsinki, Finland
CHRISTIAN GUCKELSBERGER, Department of Computer Science, Aalto University, Espoo, Finland and School of Electrical Engineering and Computer Science, Queen Mary University of London, London, UK

While creative artificial intelligence (AI) is becoming integral to our lives, we know little about what makes us call AI "creative". Informed by prior theoretical and empirical work, we investigate how perceiving evidence of a creative act beyond the final product affects our assessment of robot creativity. We study embodiment morphology as a potential moderator of this relationship, informing a 3×2 factorial design. In two lab experiments on visual art, participants ($N = 30 + 60$) assessed drawings produced by two physical robots with different morphologies, under exposure to product, process and producer as three levels of perceptual evidence. The data supports that the human assessment of robot creativity is significantly higher the more is revealed beyond the product about the creation process, and eventually the producer. We find no significant effects of embodiment morphology, contrasting existing hypotheses and offering a more detailed understanding for future work. The latter is also informed by additional exploratory analyses revealing factors potentially influencing creativity assessments, including perceived robot likeability and participants' experience with robotics and AI. Our insights empirically ground existing design patterns, foster fairness and validity in system comparisons, and contribute to a deeper understanding of our relationship with creative AI and thus its adoption in society.

This work was financially supported by the Academy of Finland (#328729, CACDAR) and the Helsinki Institute for Information Technology.
Authors' Contact Information: Niki Pennanen (corresponding author), Department of Computer Science, Aalto University, Espoo, Finland and Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland; e-mail: niki.pennanen@aalto.fi; Simo Linkola, Department of Computer Science, University of Helsinki, Helsinki, Finland; Anna Kantosalo, University of Helsinki, Helsinki, Finland; Nicolas Hiillos, Department of Computer Science, University of Helsinki, Helsinki, Finland; Tomi Männistö, Department of Computer Science, University of Helsinki, Helsinki, Finland; e-mail: tomi.mannisto@helsinki.fi; Christian Guckelsberger, Department of Computer Science, Aalto University, Espoo, Finland and School of Electrical Engineering and Computer Science, Queen Mary University of London, London, UK; e-mail: christian.guckelsberger@aalto.fi.

ACM 2573-9522/2025/4-ART41
https://doi.org/10.1145/3711939

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; User studies; Laboratory experiments; • **Applied computing** → *Fine arts*; *Psychology*;

Additional Key Words and Phrases: Human-Robot Interaction, Computational Creativity, Creative AI, Human Perception, Embodiment

ACM Reference format:
Niki Pennanen, Simo Linkola, Anna Kantosalo, Nicolas Hiillos, Tomi Männistö, and Christian Guckelsberger. 2025. From Product to Producer: The Impact of Perceptual Evidence and Robot Embodiment on the Human Assessment of AI Creativity. *ACM Trans. Hum.-Robot Interact.* 14, 3, Article 41 (April 2025), 41 pages.
https://doi.org/10.1145/3711939

1 Introduction

For many decades [10], computational creativity[1] researchers have sought to understand whether, and how, we could engineer **Artificial Intelligence (AI)** systems which can be considered creative in their own right [29]. New generative machine learning architectures such as transformers [116] and diffusion models [54] have recently advanced this agenda at an unprecedented scale, forming the foundation of systems capable of producing high-quality artefacts in many artistic domains, from visual and concept art, over music and video, to literary fiction [38]. There exists consistent evidence that people experience emotional responses to AI-generated artwork and attribute intentions to it [33]. Over the past 2 years, we witnessed a rapid adoption of such "creative AI"[2] by human creatives [57, 67, 75, 117]. The vast majority of these systems access our world indirectly through data collected by people and act on it through the impressions they leave on their users [82]. Already soon, we expect more such creative AI to become equipped with a body comprising sensors and actuators, and interact with us in the same, shared physical space. Such *embodied computational creativity* systems are by no means dreams of the future; Guckelsberger et al. [46] have identified a surge in artistic and engineering research activity over the past 10 years bringing forth embodied creative systems across many domains, including drawing and painting robots [49], robot musicians [13] and performative artists [42, 43].

Many of the big questions raised by this development [38] are of immediate relevance for **Human–Computer Interaction (HCI)** researchers and Psychologists interested in creativity: How will AI change future creative practice, and how will it be adopted by creatives in their work [75, 117]? How will AI change our very understanding of creativity [86]? Who will receive credit for AI-generated artefacts [39]? How can we enhance people's trust in creative AI systems and their creators [91]? And, if AI was ever autonomously creative, would we recognise and acknowledge it as such [47]? Highlighting more of these questions, Epstein et al. [38] argue that "understanding the impact of generative AI—and making policy decisions around it—requires new interdisciplinary scientific inquiry into culture, economics, law, algorithms and the interaction of technology and creativity". Our focus here is on the latter; we hold that any inquiry into these questions would benefit from a better understanding of the factors that make us call AI "creative" in the first place, i.e., which shape our subjective perception of AI creativity or what has recently been coined *the Lovelace effect* in honour of Augusta Ada King, Countess of Lovelace [91].

[1] We use the terms "creative AI", "artificial creativity", "machine creativity" and "computational creativity" interchangeably; while Computational Creativity can be considered a wider research programme, the relationship of these terms is a matter of ongoing debate.

[2] In this article, we are not concerned whether AI is or can ever be truly creative in its own right and independently of an observer's assessment [47, 58]; instead, we focus on better understanding people's perception of creativity in AI [91], and interaction effects arising from individual differences.

Crucially though, while people's perception of creativity in machines has been of interest to researchers from many fields for a long time and many theoretical frameworks for the evaluation of computational creativity have been proposed (e.g., [25, 58, 78]), researchers lament that solid empirical studies on people's perception of creative AI, informed by Psychology and HCI, are still rare [19, 73, 91]: "A key problem, indeed, is that we do not yet have enough understanding of the circumstances that lead users to attribute creativity to machines" [91]. In the present study, we investigate these circumstances, focusing on the impact of ***Perceptual Evidence (PE)*** and *embodiment* on the human attribution of creativity to machines. We motivate the study of both variables based on theory and empirical evidence from Creativity Studies, Computational Creativity and **Human–Robot Interaction (HRI)** [45] research.

1.1 The Role of PE in the Assessment of Creativity

We investigate *PE* for its deep connection with the four perspectives (or "four P's") on creativity [59, 85, 100], one of the most important metatheories of creativity [44] which synthesises our understanding of creativity into four perspectives: the creative product (artefact), process (act), producer (person/machine) and the press, the sociocultural environment in which creativity unfolds. These in turn fundamentally shape future creativity theory and empirical work [101]. While many studies look at manipulating aspects of the same perspective, e.g., revealing more about the creative process and demonstrating an effect on people's creativity assessment [69], only few researchers interpreted the perspectives as different degrees of PE of the creative act [51, 78], and even fewer empirically compared the effect of such perceptual exposure on the assessment of creativity [19]. In other words, the vast majority of research on the human perception of creativity is "within-P", rather than "between-P". This is surprising, given that the attribute "creative" can be applied to three of these perspectives at once: producer, process and product, and, *vice versa*, the most popular definitions of creativity feature multiple of these perspectives. To us, this suggests that exposure to these three P's alone may shape our attribution of creativity. Based on this observation and the earlier mentioned studies, we hypothesise that:

H1: Exposure to the product, process and producer as PE of the creative act causes different assessments of creativity by a separate observer.

1.2 The Role of Embodiment in the Assessment of Creativity

As our second variable, we investigate the effect of *embodiment* as a feature of the producer. Our motivation here is threefold. Firstly, the PE interpretation of the four P's on creativity has featured prominently in research proposals [46, 53], theoretical frameworks [78] and early empirical studies of embodiment effects in creativity [52, 87], implicitly suggesting an interaction effect. Such an interaction is explicitly supported by theoretical and empirical HRI research, identifying people's perception of an artificial entity's body-related capabilities as mediator of embodiment effects [55]. Secondly, HRI has not studied the impact of embodiment effects on creativity, despite an increased perception of creativity having been connected with stronger feelings of trust [91], which is of core interest to HRI researchers. Including this facet thus enables us to close a critical gap in HRI research. Finally, most existing research focuses on manipulations of the other perspectives but the producer; thus, considering variations of the producer provides more robustness to our study of PE against the backdrop of existing research. Existing work has investigated embodiment effects mostly on variables that are adjacent to, or constituting [60, 102], the human perception of creativity, e.g., people's aesthetic assessment [18, 19] or perceived empathy [93, 108]. The few studies assessing creativity have either not differentiated the producer [19], considering a system's physical presence, embodiment, interaction and environment all at once, or have not investigated how variations in these factors affect the creativity assessment [87]. Moreover, these studies focus on the existence or absence of embodiment, but do not address how differences within an artificial producer's embodiment affect people's assessment of creativity. Here, we focus on such differences in the morphology of two physically embodied systems, i.e., robots. We focus on morphology, as it has been proposed as a central mediator of embodiment effects in HRI [55] and supported by a small body of very recent research [70, 103, 113] for other outcome variables but creativity. Human-like morphology has also been identified as one of the key traits that supports coordination with and social inclusion of artificial creative agents [23]. Moreover, proximity to our human morphology is usually assumed as a given in creativity and empirical aesthetics research, but has been emphasised as one of the most distinguishing factors between human and machine embodiment in related computational creativity work [28, 46, 47, 83], prompting further inquiry. Informed by this research, we assume that the role of morphology in embodiment effects also translates to people's attribution of creativity to AI, and hypothesise:

> *H2*: Different physical (robot) embodiments, distinguished by morphology, cause differences in people's assessment of creativity, but only if perceiving product, process and producer together.

1.3 Research Contributions and Methods Overview

We contribute the first quantitative study that systematically investigates the effect of perceiving product, process and producer on the human assessment of AI creativity under controlled lab conditions. Moreover, we are first to investigate the effect of morphology as feature of physically embodied producers on people's creativity assessment.

We assess our hypotheses through a within-participants study (N = 60, Section 3), for which the materials were tested and pre-selected in a smaller pre-study (N = 30, Appendix B). We support comparisons to related work by focusing on visual art, more specifically drawing. We inform our study design by existing research [78], factorising what is commonly denoted the "perception of creativity" [24] into participants' assessment of Creativity[3] and PE. The latter denotes the observable aspects of the system being assessed and is operationalised as three of the four perspectives on creativity [58, 85, 100]: Product, Process and Producer. PE is complemented by the system's Embodiment as a second independent variable. We compare two physically embodied systems, i.e. robots, differentiated in terms of mechanistic vs. organismoid morphology [121] and corresponding to a plotter and a robot arm, respectively. We employ a factorial design, which involves simultaneous manipulation of the two independent variables. This approach allows for the investigation of interaction effects between variables, in addition to the main effects. Using this design, we are able to identify for instance potential unwanted differences in the creativity assessment of processes or products caused by different embodiment morphologies.

As part of our complementing exploratory analysis, we investigate previously underexplored interaction effects between robotics and AI expertise and people's assessment of creativity. For the measurement of our dependent variable, we adopt a non-ontological view of creativity with a long tradition in Psychology (summarised by Amabile [5]) and Computational Creativity [29] as the "reactions and perceptions of human users who enter into interactions with" artificial systems [91]. We match this understanding methodologically by adapting the well-established **Consensual Assessment Technique (CAT)** [4] to the HRI domain. We hold the creative process exercised by our robots close to constant, thus isolating the effect of embodiment on our perception of creativity from its effect on creativity itself. This is not a contradiction with, but direct consequence acknowledging the role of embodiment in shaping creative cognition and action. These methodological decisions

[3]We distinguish experimental variables and their levels through capitalisation here and in the study sections.

allow us to further advance existing work by assessing creativity as a multifaceted construct and distinguishing embodiment morphology from related factors such as physical presence and embodied interaction.

1.4 Findings and Implications

Our results indicate that the human assessment of robot creativity is significantly influenced by PE, supporting our first hypothesis. Specifically, we found that as more information about the creative act was revealed, from the product, to the process and finally to the producer, creativity assessments increased at each stage. Surprisingly, we observed no effect of embodiment morphology on creativity assessments, contradicting our second hypothesis and previous literature. Additionally, our exploratory analyses identified several factors that may also affect creativity assessments, including perceived robot likeability and participants' experience with robotics and AI.

This research has implications (cf. Section 6) for design, public awareness and scientific progress. In existing design practice, PE, especially w.r.t. facets of embodiment, is deliberately injected into what Colton et al. [27] call "computational creativity theatre" to heighten people's perception of creativity in artificial systems. Examples range from portrait painting software that renders a human hand mimicking the process of painting [26], over robot musicians enacting their performance [107], to AlphaGo's match against Lee Sedol, in which the software's moves were physically performed by a person on a material board [12, 91]. With the present study, we aim to provide empirical support for these existing design patterns, and, *vice versa*, foster awareness in society of how our assessment of creativity may be manipulated through such design decisions. On the scientific side, our findings can inform how artificial systems should be compared with respect to their creativity in future study design, support the critical assessment of existing studies w.r.t. the validity of the reported findings and inspire further research questions on our constructs and the Lovelace effect more generally.

We limit the scope of this work to the observation of an "autonomously creative" system, as opposed to the rich interaction with a co-creative system. This is a deliberate decision (cf. Section 3) to eliminate further confounders, and we consider this article a necessary stepping stone toward the study of PE and embodiment effects in rich human–machine co-creative interaction. We project implications of our findings on co-creativity in Section 6. We moreover understand creative AI as embodied, encompassing not just data and algorithms, but also the system's body, interactions, perceptions and contextual influences that holistically shape its creative expression. We argue that, as long as creative AI algorithms are not evaluated in isolation but presented theatrically [27], or make actual use of a system's embodiment, the effect of these additional constituents must be taken into account. While our study does not directly assess features of AI models, it yet contributes to creative AI research because our findings were obtained from participants evaluating stimuli under the assumption (see Section 4 for discussion on the effectiveness of the deception) that they were observing an embodied AI system. Consequently, our results are inherently influenced by this belief in AI, and our study provides implications for understanding and planning future research in creative AI.

2 Background and Related Work

In the present study, we focus on the impact of PE and embodiment morphology on the Lovelace effect. We next introduce background and related work to motivate each of these two variables, our hypotheses and our experimental design, and highlight the novelty and significance of our study relative to existing work.

2.1 PE: Product, Process and Producer

We inform PE as the first factor in our experiment by "probably the most often-used structure for creativity studies" [101] and "part of the canonical body of theories in the creativity literature"

[44]: the *four perspectives on creativity*, commonly abbreviated as the *four P's*. Identified seemingly independently by Rhodes [100] and Mooney [85], this meta-framework represents one of the major efforts in disambiguating and synthesising existing constructs of creativity. Comparing more than 40 definitions of creativity, Rhodes [100] concluded that creativity theory reflects four distinct perspectives: the *person* is the originator of the creative work, e.g., an artist. Studies from this perspective focus on which of this agent's features contribute to them being (perceived as) creative [73]. The *process* denotes the internal and external steps an individual undertakes when being creative, e.g., drawing lines on paper. The *product* is the result of this process, in our case a drawing. Finally, the *press* relates to the sociocultural environment which shapes our views on, and our assessment of, creativity. As a metatheory, the four P's structure perspective-specific theories of creativity, e.g., Wallas' [119] four stages of the creative process, or Guilford's [48] distinction of convergent and divergent thinking in the producer. The four P's have not only been important as a synthesis of existing theory, but became influential in "shaping creativity as an emerging academic discipline" [44], structuring creativity assessment methodology (e.g., [2, 9]), reviews (e.g., [41, 73]) and empirical research (for an overview, cf. [44]) in Psychology, HCI and other domains. Unsurprisingly, the four P's also feature prominently in the most common meta-definitions of creativity. For instance, Puryear and Lamb [98] as well as Plucker et al. [96] conceive creativity in the interaction of producer, process and environment, and Walia [118] emphasises the need to provide "insights into the creative act (i.e., process) itself, a factor that has been neglected in earlier definitions of creativity" [118]. Finally, the four P's have been adapted and extended in many ways, "rewriting the language of creativity" by reflecting different or new psychological traditions. We particularly adopt (more detail below) Glăveanu's [44] five A's of creativity, which draw on Sociocultural Psychology, models of the distributed and extended mind, and Ecological Psychology to offer a "cultured" or "socialised" version of the original four P's. Also important for the present study, the framework's relevance stretches beyond human creativity into the computational domain: Jordanous [59] has translated the four P's to Computational Creativity [29] more generally, and Kantosalo and Takala [63] to human–machine co-creativity specifically.

Given the frameworks' origins as means of synthesis, it is unsurprising that several authors pointed out an anomaly in the use of "creative", in that the attribute can be applied to the three perspectives of human/artificial producer, process and product individually, but also to the overall system of all four P's [59, 112]. This informs our first hypothesis (*H1*, Section 1): given that we can attribute creativity to each perspective, we hypothesise that the perception of creativity differs with the inclusion of different perspectives. The present study focuses on product, process and producer only, accumulated in a factor denoting the *PE* an observer witnesses about the creative act exhibited by an artificial system. The values are constituted by the product only (the drawing), the product and process (the drawing while being made) and all three together (the robot in the act of creating the drawing). It is also because of this anomaly that we refer to everything the participant perceives (rather than a specific perspective) as object of the creativity assessment in our questionnaire across conditions (cf. Section 3.3).

While relating to the three P's in the classic notation, we crucially adopt important semantic differences from two extensions. As proposed in Glăveanu's [44] five A's framework, we understand the process as embedded within the broader concept of action, thus not only acknowledging an internal but also an external, behavioural expression, the latter which is under investigation here. We moreover adopt Jordanous' [59] translation of the 4P's into the computational domain, in that we substitute *person* with the term *producer* to accommodate artificial agents.

As summarised by Glăveanu [44], the four perspectives on creativity are omnipresent in both creativity theory and empirical studies. The *product* features prominently in empirical studies of creativity, such as Amabile's [3] classic studies of reward effects on creativity. Moruzzi [88] has conducted a factorial study [an extension of 87, cf. below] with imaginative scenarios blending only the process and producer of an artwork. Qualitative feedback from several participants highlights the importance of the product in the assessment of creativity: "I really don't think I can answer any of the below questions (...) without having actually seen the painting" [88]. The importance of the *process* has, amongst others, been emphasised by art philosophers, arguing that people assess art holistically as the "end point of a performance" [36]. This informs the claim that information about, or direct observation of, the process influences our assessment of artistic and aesthetic value [37]. Computational creativity theorists have translated this claim to people's perception of creative AI [24]. Kruger et al. [69] provide further empirical support for the importance of information about the process through their "effort heuristic" in art: they found that participants rated the quality, value and liking for a range of artefacts, including two paintings, higher if they were informed that the artefacts took more time and effort to produce. Similarly, Jucker et al. [62] showed that higher levels of perceived investment from artists as *producers* increased participants' ratings of artistic merit. This theoretical and empirical work highlights the importance of the product, process and producer perspectives as empirical evidence for the creativity assessment and further motivates *H1*. To the best of our knowledge, no existing study has compared all three perspectives as done here. Closest to such a comparison is the work of Hennessey [51] on evaluating the reliability of the CAT beyond the product perspective. Similar to our study design, they exposed participants to visual art products and a recording of their creation process. Also similarly (and further justified in Section 3.1), participants were exposed to the product first, followed by the product's creation process. Hennessey's study demonstrates the reliability of the CAT across perspectives, which supports our choice of the same assessment method (Section 3.3). Crucially though, Hennessey only reports correlations between the ratings of product and process, but not the rating magnitudes required for our hypotheses.

2.2 Embodiment

Studies from the producer perspective focus on which agent features contribute to them being (perceived as) creative [73]. In this article, we investigate embodiment (definition below) as such a feature. Differences in measurements due to different embodiments, i.e., *embodiment effects*, are prominently investigated in the fields of Human–Agent and HRI. Our construct of embodiment can be grounded in the EmCorp framework [55], an empirically supported HRI framework, which posits that embodiment effects are mediated by people's perception of an artificial entity's body-related capabilities. This matches our focus on *perceived* creativity in the Lovelace effect [91] and our interest in different levels of PE. Next to perceived capabilities as mediators, Hoffmann et al.'s framework distinguishes several moderators of embodiment effects. Most notable for us, they argue that different *morphologies* can affect people's assessment even when perceived within the same, e.g., physical, embodiment (rather than being perceived between e.g. virtually and physically embodied agents). This motivates our second hypothesis (*H2*, Section 1) from the HRI side. We investigate the impact of morphology by contrasting mechanistic and organismoid[4] morphologies for agents of the same, physical embodiment.[5] Informed by empirical findings (see below), we project that our choice of morphologies will affect people's perception of creativity especially *via* the mediator of perceived *nonverbal expressiveness*, operationalised by questionnaire items such as "is unrestricted in its actions/movements/gestures", etc. We particularly suspect that people's assessment of machine creativity is positively affected by reducing the gap in the morphology and thus expressiveness

[4]Term adopted from Ziemke [121], corresponding to "zoomorphic" in the work of Hoffmann et al. [55].
[5]As a subtle point, Hoffmann et al. [55] consider embodiment (physical, virtual) separately from morphology. We in contrast only consider physical embodiment and, adopting the typology by Ziemke [121], write about differences in morphology as differences in embodiment.

between the robot and human observer, as supported by work in embodied aesthetics referenced below and hypothesised in non-empirical computational creativity research [28, 46, 47, 83].

2.2.1 Embodiment and Morphology in HRI. We first consider related work in HRI more generally, before discussing theoretical and empirical work on embodiment and creativity specifically. Most HRI studies on embodiment effects focus on comparisons of physical and virtual embodiment, assessing a broad range of outcomes such as experiential variables like enjoyment, immersion, perceived control and the perception of anthropomorphic traits like trustworthiness, dominance and social presence (cf. [55], for a brief review). Crucially, Hoffmann et al. attest that morphology has been widely neglected and, as of 2018, "systematical comparisons of the impact of morphologies within one physical embodiment are missing". We thus advance a very small body of research comparing morphologies within the same physical embodiment. Stroessner and Benitez [113] compare people's attribution of **Robot Social Attribute Scale (RoSAS)** [17] items and evaluative responses such as liking and desire for contact when perceiving pictures of humanoid vs. machine-like robots of different genders in an online study. Focusing on the effects of morphology only, the humanoid robots were considered warmer and more competent, and these RoSAS items worked as predictors for the evaluative responses. Comparing people's perception when talking with three robots of different morphology (humanoid, organismoid, functional) in a between-subjects lab experiment (N = 95), Sasser et al. [103] find significant differences in the Godspeed questionnaire [7] items of Anthropomorphism and Animacy. Additional exploratory analyses identified, amongst others, the EmCorp mediator of nonverbal expressiveness as a predictor of these differences. Most recently, Kunold et al. [70] have compared people's responses to images of 46 robots with different morphologies (anthropomorphic, zoomorphic, caricatured, functional) in an online study (N = 673), measuring responses on the EmCorp scale [55] and RoSAS [17], amongst others. They find that all EmCorp subscales show medium to high correlations with physical human-likeness, indicating that greater physical human-likeness is associated with higher capability perception. Through a second online study (N = 586), they confirm that robots with similar morphological features differ significantly in body-related capabilities as measured by the EmCorp-Scale, and mediation analyses revealed that these perceived capabilities further explain the robots' assessments in terms of acceptance and RoSAS attributes. They conclude that these findings further support the EmCorp framework, in that "the assessment of robots can (to some extent) be explained through the expected capabilities humans attribute to them based on their morphological appearance" [70]. These studies support our focus on manipulating robot morphology as independent variable (Section 3.1). Crucially though, no related HRI study investigates creativity as a dependent variable, and most only use *images* of embodied systems; we in contrast expose participants to two actual robots as we suspect this to affect the measurement (further motivated below).

2.2.2 Embodiment in Creativity Research. While there exists some research on the effect of producer characteristics on (AI) creativity, research on embodiment effects is very scarce. Chamberlain et al. [19], connecting work from Cognitive Neuroscience and Psychology, hypothesise an effect of physical embodiment on the human aesthetic assessment via mirror neuron activity. This is supported by existing studies on human creativity, demonstrating that traces of embodied action encoded in human art can evoke, amongst others, mirror neuron responses with an affect on the aesthetic assessment [18, 40, 115]. Chamberlain et al. [19] further support the connection of embodiment and aesthetic assessment by drawing on artist's experience with creative robots, highlighting that "working under the constraints of the physical work (...) is arguably more challenging than producing e.g. an image on a virtual canvas". Crucially, this selection of work focuses on the aesthetic assessment as constituent of creativity, but not creativity per se. Following a survey of embodied computational creativity work, Guckelsberger et al. [46] argue for the importance of understanding the effect of embodiment on people's perception of AI creativity to support system evaluation and design. Linkola et al. [78] propose a framework for investigating embodiment effects on people's perception of creativity, and put forward concrete study proposals, including comparisons based on differences in morphology, and distinguishing PE by means of the 4P's. We reflect this proposal in our hypotheses *H1* and *H2* (Section 1). Resting on this work, Herman and Moruzzi [53] hypothesise a positive correlation "between the attribution of creativity and the embodied presence of the actor performing the process under examination", but a complete analysis of findings from their mixed-method study is pending. Herman and Hwang [52] complement the earlier studies on aesthetics with qualitative findings from a longitudinal online study, in which participants attributed more creative value to man-made visual art if it showed traces of the human embodied process, such as brushstrokes. Moreover, Moruzzi [87], studying creative collaboration, puts forward qualitative findings suggesting that, when a person is presented "together with an embodied artificial agent (...), the rating of creativity and perception of collaboration is increased" but, "when the artificial actor is not embodied, the creativity, when at all recognized, is attributed to the human actor alone". The latter highlights the importance of investigating the role of embodiment for better understanding the perception of human–machine co-creative partnership [65]. This selection of theoretical and empirical work not only demonstrates the importance of the product, process and producer, but also highlights ambiguity in how embodiment as a characteristic of the producer is conceived and investigated: the mere presence of an embodied agent, embodied interaction witnessed either explicitly or implicitly through the artwork or, as in this study, the producer's body and sensorimotor equipment. This ambiguity further motivates the clean distinction of perspectives which we seek to investigate for our first hypothesis *H1*. Crucially, all quantitative insights in related work concern the effect of PE on people's aesthetic assessment. We next provide more detail on the few studies addressing effects on creativity, as most relevant for this work.

Moruzzi [87] has conducted a factorial survey study [6] to investigate the influence of embodiment and other factors on people's evaluation of human and AI creativity. Participants were presented with vignettes of hypothetical scenarios, varied by the factor levels. In a mixed design online, each participant (N = 161) was asked to rate their perceived creativity given two text descriptions, of which one focused on painting. Different participants received descriptions of the painting actor being either a software, or embodied in the form of a robot arm. Moruzzi finds that the embodied actor is perceived as less creative than their software counterpart, but without statistical significance. The differences to the present work are threefold. (1) The factorial survey design required participants to imagine, rather than directly perceive, the process described in the vignettes. We hold that a scenario imagined from a text description will evoke a less intense experience of embodiment compared to witnessing it in the same physical space. For improved validity, we exposed participants to real artefacts and embodied systems in the same physical space. (2) Moruzzi acknowledges to only contrast the presence and absence of physical embodiment, while we investigate the effect of two morphologies of physically embodied robots. (3) Finally, Moruzzi's study focuses on descriptions of a single process and variations in its producer, but does not compare variations in either the process or product. We compare perceptions of product, process and producer on different levels for product and producer.

Chamberlain et al. [19] investigated factors shaping human biases against AI art as expressed through the people's aesthetic assessment. In a between-groups design across two sites, they measured responses to portraits made by robots comprised in an art installation by Tresset and Leymarie [114]. In the first condition (n = 145), set in a fine art space, participants could witness the robots in action and interact with them to have their own portrait drawn. In the second condition (n = 97), set in a university library, participants examined a display of drawings only by the same robot, being made aware of their originator.[6] In both conditions, participants responded to questions on perceived value and aesthetics, and their art interest. In the interactive condition, they additionally received questions pertaining to creativity and the perception of robot attributes. Most relevant here, Chamberlain et al. find that participants rated the creativity of drawings higher in the second condition, with strong statistical significance. The present work complements and extends this study in several respects: (1) Chamberlain et al. are primarily concerned with differences in people's perception of human and robot artists, while we focus on differences between machines only. (2) Their study focuses on the aesthetic assessment, while we focus on creativity. We assess creativity throughout all conditions, thus enabling comparisons. (3) Chamberlain et al. do not isolate the producer from the process; we in contrast systematically distinguishing all three levels of PE. (4) Their understanding of embodiment is wider than ours, in that it also comprises interaction with the participants. We crucially do not involve the participant as artistic subject to avoid additional noise in the assessment, and to reproduce a situation closer to creative partnerships. We invested substantial effort to keep the process between robots and repeated drawings similar, setting up robots and drawing routines specifically for this study. (5) We compare two robot embodiments differing in the robot morphology (mechanistic, organismoid) as a separate factor and potential moderator, responding to the HRI call for more comparative studies on embodiment morphology. (6) A limitation well acknowledged, not testing participants in the same setting essentially means to not control for the Press, the sociocultural environment in which creativity happens (cf. Section 2.1). Recent critical inquiry has highlighted the potential of art spaces to change our perception of machine-made art [91], an effect well known for man-made art. We consequently conduct a within-participant design in the lab, counteracting participant selection bias and recruiting a more heterogeneous demographic. Finally (7), we also record robotics expertise, allowing us to identify interaction effects with the assessment of creativity (Section 3.6). Together, this theoretical and empirical work in creativity studies support our focus on investigating embodiment effects in creativity w.r.t. the impact of morphology, as expressed in *H2*.

2.3 Interaction Effects

We study PE and embodiment in a factorial design to assess interaction effects between these variables. Such interactions have been supported by work in HRI such as the EmCorp framework of Hoffmann et al. [55]. In emphasising people's perception of an agent's body-related capabilities as mediator in embodiment effects, it supports that the extent to which people perceiving body-related capabilities, as captured in our notion of PE and differentiated by the three P's, should be considered as a separate, interacting variable. Moreover, studying these interactions allows us to provide stronger support for the robustness of embodiment and PE effects, given different types of PE and embodiments, respectively. Furthermore, research in the domain of AI-composed music indicates that although the anthropomorphic characteristics of an AI composer might not influence evaluations of the music (product) itself, they can shape perceptions of the musician (producer), potentially enhancing the acceptance of the system in a musical role [56]. To determine if similar effects exist for different artistic domains such as visual art in our study, it is essential to employ factorial study designs.

This work is interdisciplinary in that it combines research questions, insights and methodology from related work in HRI and Human–Agent Interaction with creativity-related sub-disciplines of Psychology and AI in order to study people's perception of creative machines. Our survey of

[6]Chamberlain et al. [19] also investigated a third condition which was otherwise identical, but participants were not told about the identity of the artist. We do not discuss results related to this condition, as it is not immediately relevant for this work.

related work in this section highlights that, while promoting studies on embodiment effects, the role of morphology (*H2*, Section 1) has not been addressed in disciplines interested in creativity, but such studies are strongly advocated by HRI researchers. *Vice versa*, while acknowledging the role of people's perception for embodiment effects and calling for further investigation, HRI research has so far not addressed the impact of PE distinguished by the different perspectives (*H1*, Section 1) which are commonplace in the creativity-related sub-disciplines. Where possible, we compare our findings to this related work in Section 4.

3 Study

In order to answer our research questions (Section 1), we carried out a within-subjects lab experiment where participants assessed the Creativity of drawings, under the assumption that an autonomous robot had created them. As noted earlier (Section 1), we adopt our experimental design from Linkola et al. [78] who factorise the perception of creativity into (1) PE, limiting the visible features of the system and its behaviour, and (2) the assessment of the embodied system's creativity based on the available PE. This PE is assumed to be shaped by (3) the system's embodiment (including morphologies as described in Section 2.2), and (4) the creative process that it executes. We consequently aimed to keep the system's creative process as constant as possible, and measured the participants' perceived creativity assessments as our dependent variable. Our experimental task focused solely on observation and did not include direct interaction with the system. As Hoffmann et al. [55] emphasises, different interaction scenarios can make certain aspects of morphologies more salient. Our objective was to examine the subject from a more neutral standpoint, without the moderating effect of an interactive task.

The study and specifically its deceptive element has been approved by the Aalto University's Research Ethics Committee. The study was preregistered on the **Open Science Framework (OSF)** prior to data collection. The data used for analyses and other supplementary materials can be found in the OSF Supplementary.

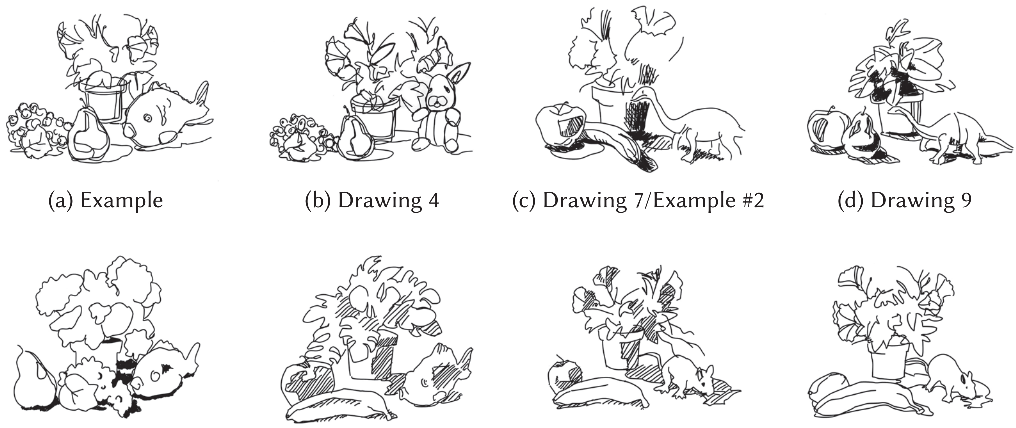
3.1 Study Design

We adopted a within-subjects study design to maximise statistical power given our sample size, while also aiming to minimise error variance due to individual differences. We employed a full factorial design (3×2), participants were exposed to all combinations of independent variables. The two manipulated independent variables were:

PE. Which aspects of the systems and its behaviour the participants can perceive. The levels included (1) the final Product only, (2) both the final Product and the creation Process and (3) the final Product, Process and the Producer (robot). These stages accumulate what has been considered different perspectives on creativity [59, 100].

Robot Embodiment. Which robot created the drawing, the levels being (1) AxiDraw (mechanistic morphology) and (2) xArm (organismoid morphology). More details on the robots in Section 3.2.

To mitigate order effects, the sequence of drawings and Embodiments was counterbalanced. However, since the PE levels are accumulative, we presented them in a fixed order for all participants. This approach is similar to the one in [51] and was chosen to avoid leaking impressions from a richer level of PE into another, potentially biasing people's assessment. In a counterbalanced order, the exposure to higher levels of PE (such as Process or Producer) before the Product-only condition could influence participants' perception, which are meant to be isolated to the Product. Therefore, we anticipated that any bias introduced by a fixed order would be less significant than the bias from



(a) Example (b) Drawing 4 (c) Drawing 7/Example #2 (d) Drawing 9

(e) Drawing 10 (f) Drawing 11 (g) Drawing 14 (h) Drawing 16

Fig. 1. Selection of drawings for the study, informed by the pre-study results (Appendix B.5).

leaking evidence from varying orders. We address the limitations of this approach and alternatives for future work in Sections 5 and 7.

3.2 Materials

Drawings. As stimuli for participants to assess in the study, we used still life drawings of object arrangements, each consisting of one artificial plant, one toy animal and two artificial fruits. We chose still life drawings due to their prominence in art history and common artistic training, ensuring familiarity to many participants and thereby minimising distractions. Focusing on drawings also supports direct comparisons to existing work. With a long history, drawing and painting are arguably the most popular artistic domains that have been transferred into embodied systems to date. Furthermore, these simple arrangements include only few objects, with no moving or living parts. This allows for consistent reproduction and quick sketching, fitting well for the time limitations of our study. The drawings and object arrangements are described in more detail in Appendix B.1.

Eight drawings were used as stimuli in the experiment (see Figure 1). To ensure that our drawings had the potential to be considered creative, we conducted a pre-study, where 30 participants assessed 18 candidate drawings (see Appendix B). Based on the results of the pre-study (Appendix B.5), we selected seven drawings close to each other in assessed Creativity, and the drawing used as an example drawing in the pre-study, to be used in the study. Of the seven, the drawing with median Creativity was chosen as a second example drawing (Drawing 7), while the other six acted as the main stimuli to be assessed. The two examples served as practice stimuli to explain the task, familiarise participants with the still life subject, and provide them with a baseline for the assessment of the remaining drawings. We introduced the second example drawing in response to pre-study feedback, where participants indicated difficulty in rating the initial drawings without adequate baseline.

Robots. To contrast very different physical embodiments [121] and to reflect some of the variety in state-of-the-art drawing and painting robots in academic and artistic contexts [49], we recreated the drawings with the following two robots: Evil Mad Scientist AxiDraw V3/A3 and UFactory xArm Lite 6. Figure 2 contains images of both robots. Videos of the robots drawing can be found in the OSF Supplementary.

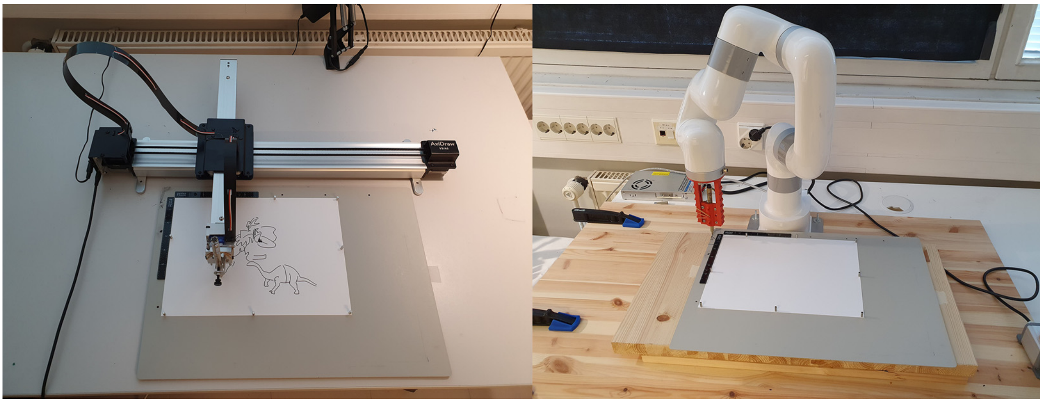
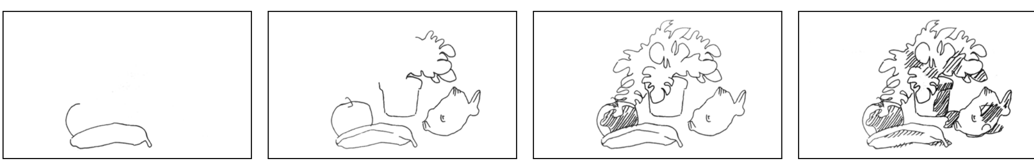


Fig. 2. Robots used in the experiment: Evil Mad Scientist AxiDraw V3/A3 (left) and UFactory xArm Lite 6.



(a) 0:11 (b) 01:09 (c) 02:32 (d) 04:58

Fig. 3. Example still images from a randomly selected video used as a stimulus in the Product+Process condition.

AxiDraw is a robotic pen plotter, commercially produced for drawing and writing purposes. We include it for its simple, mechanistic and transparent embodiment with three degrees of freedom. In contrast, xArm is a general-purpose shiny white robotic arm with metallic colouring near the joints. We included it for its complex embodiment with six degrees of freedom, mimicking limbs as observed in the animal kingdom. Together, these robots allowed us to investigate the differences caused by mechanistic and organismoid [121] morphologies, as discussed in Section 2.2. They are described in more detail in Appendix C.

Videos. For the PE level Product+Process, we utilised videos showing the drawing process of the robots as stimuli, in addition to the drawings. These videos were recorded from below a transparent table, ensuring that the embodiment of the robot was not visible. Example still images from a video are presented in Figure 3, and the full videos are available in the OSF Supplementary.

3.3 Questionnaires

We administered three different types of questionnaires during the experiment.

Demographics and Expertise Questionnaires. We gathered demographic information and assessed participants' experience with art, AI and robotics through an electronic questionnaire (OSF Supplementary). Demographics comprise age, identified gender, level of general education and information about art education. Art interest was measured with the Art Interest sub-scale of the validated [109] **Vienna Art Interest and Art Knowledge (VAIAK)** questionnaire [110]. Since this questionnaire does not distinguish art experience through spectatorship and practice, and we expected artists amongst the participants, we added a question on how often they practised visual arts.

At the time of preparing this experiment, we could not find any published, validated and commonly used questionnaires to measure our participants' familiarity with AI and robotics. We consequently decided to adapt an existing AI literacy scale [16, 95, 120]. We chose to adopt the AI Technology Knowledge and AI Usage Experience sub-scales from the questionnaire by Pinski and Benlian [95]. This scale was chosen based on its status as a validated universal measure of AI literacy, unaffected by specific human roles. To assess familiarity with robotics, we added two statements using the same scale as in the AI literacy questionnaire, 7-point Likert: "Robotics is a familiar topic to me" and "I have personal experience of using robots".

Creativity Questionnaire. Compromising between acknowledging the subjective and contested nature of creativity [60] and shedding more light on the underlying components that contribute to our creativity assessment, we adopt the CAT [4] for our measurement of creativity as a dependent variable, aligning our questionnaire design closely with a large body of existing studies. The important characteristic of these CAT questions is that they require participants to apply their own subjective conception of creativity, rather than one provided by the experimenters. It also matches our position that the creativity of a system should not be assessed in ontological terms but conceived through the perception and attribution of its observers, as articulated in the Lovelace effect [91].

For improved ecological validity, we deviate from the original CAT technique by not only inviting expert participants but embracing a general audience. Whether judges require expert-level domain knowledge for high reliability is still debated, with evidence supporting both perspectives [32, 66]. However, non-experts have been shown to be particularly effective judges when the tasks or artefacts are simple in design and when the judges have some familiarity with the type of object being evaluated [32]. Given the simplicity of our drawings, we believe this mitigates the need for expert raters, as most individuals are likely familiar with similar drawings inspired by classic still life arrangements.

Leveraging existing reviews [32, 79], we informed our selection of questions based on their fit to our specific study and usage frequency in previous CAT studies. Our main interest was on the perceived Creativity of the drawings; we complemented the corresponding question by further questions on Liking, Aesthetic Appeal, Originality, Technical Artistic Ability and Detail. These represent variations of novelty and value, which are widely regarded as the two core components in the assessment of creativity [102]. Including them thus promised to add more granularity to the perceived creativity assessment. All our dimensions were also used in Amabile's original CAT studies [4]. Although she did not use originality explicitly, her dimensions related to novelty were very similar.

We worded the questions as close to Amabile's formulation [5] as possible or used other wordings from previous CAT research if more suitable. Changes were only made when necessitated by our experimental design. Most importantly, we modified the object of evaluation from "the design" or "the work" to "what you observe". This was done to allow participants to relate the questions flexibly to accumulating PE, making use of the anomaly in attributing creativity introduced in Section 2.1. Similar to Amabile's original CAT studies [5], we used visual analogue scale scored from 0 to 100, with endpoints labelled as "Low" and "High". The full questionnaire and justification for any changes to earlier formulations is presented in Appendix A.

Robot Perception Questionnaire. In order to capture people's perception of the robots when exposed to the Product+Process+Producer level of PE, we complemented these measures with the Animacy, Likeability, Perceived Intelligence and Perceived Safety sub-scales from the widely used and validated Godspeed Questionnaire [7], a standard questionnaire in HRI studies.

3.4 Participants

We initially recruited 60 participants (different from our pre-study participants, Appendix B) for the experiment conducted in July and August 2023. Due to robot malfunctions that invalidated

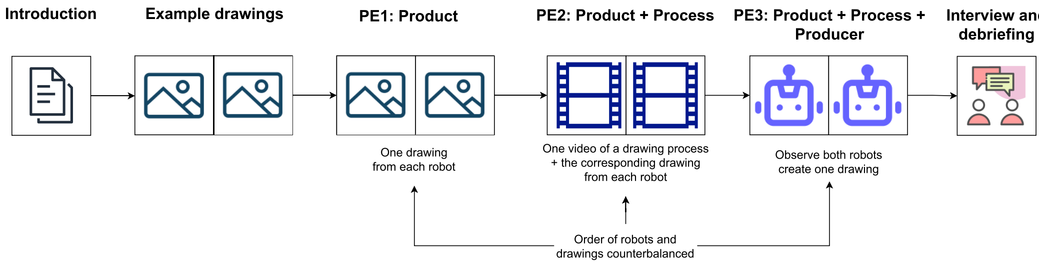


Fig. 4. Study experimental procedure.

two sessions, we recruited an additional two participants, resulting in a final sample size of 60. Participants were recruited *via* online and offline advertisements for the communities of Aalto University and University of Helsinki, and, to a much lesser extent, our personal and professional networks. Participants were required to have normal or corrected to normal vision, but this was not assessed in the study. Each participant received a 25€ gift card as compensation. Each participant gave their written informed consent, and the experiment was conducted in accordance with the Declaration of Helsinki.

Within our sample, 32 identified as females, 25 as males, one specified other, and two did not specify. Participants' ages ranged from 20 to 59 years ($M = 28.0$ $SD = 6.6$) and five participants chose not to specify their age. Eighteen participants reported having some art education (not limited to visual art), including various art domains and courses taken in childhood. Eight reported to own Doctoral, 31 Master's and 17 Bachelor's degrees or equivalent as their highest completed degree, and three reported high school level education. One participant chose not to provide this information. Summaries of interest and expertise variables as well as comparisons to pre-study participants can be found in Appendix D.

3.5 Procedure

The experiment was conducted at the premises of Aalto University and took approximately 60 to 90 minutes to complete. Participants were offered to communicate with the experimenter in either English or Finnish, depending on their preference. We provide an overview of the procedure below and in the Figure 4; the complete procedure can be found in the OSF Supplementary.

Introductions. Participants were first familiarised with the study through a written introduction, and were then asked to provide written informed consent. As a deception, they were informed that the drawings presented in the study were drawn autonomously by unspecified robots.

They were next shown the same example drawing, and a printed photo of the depicted arrangement, similarly as in the pre-study (Appendix B) to familiarise them with the task and to provide a baseline. Participants then received another drawing as a training stimulus to practice rating our questionnaire items on an iPad Pro 11. This drawing was also the same for everybody.

Product. Participants entered the main part of the study by being exposed to the first, Product-level of the PE variable. They separately rated two drawings previously created by the robots, each accompanied by a photo of the original arrangement. One drawing was made by the xArm, and the other by the AxiDraw (Section 3.2). Participants were not provided with any information about the specific robot behind the drawings.

Product+Process. Participants then viewed two separate videos of the robots' drawing processes from the iPad, one video from each robot. In other words, the PE of Product-only was augmented by its Process of creation. Participants were also given a photo of the original arrangement while

Fig. 5. Experimental setup including the robots (AxiDraw, left; xArm), arrangement, camera and participant chair. Missing from the picture is the folding screen, which was initially used to keep the robots out of sight.

watching the videos. Following each video, they received the drawing they had observed being created and were asked to rate it.

Product+Process+Producer. In the last stage of the study, participants separately observed the two robots as they create the drawings live. While the robots were present in the same room throughout the whole experiment, they have been kept out of sight for the previous levels of PE using a folding screen. For this final level, they were revealed to the participant one at a time. The room also features a previously hidden physical object arrangement and a camera positioned towards it. The whole setup (see Figure 5) was designed to mirror the original human artist's drawing setup (see Figure B2 in Appendix. B.1), with the camera positioned similarly to the artist's eyes relative to the arrangement. Light source and angle were also matched to the original conditions to keep the shadows as close to the setting which has been captured by the artist. Participants were informed that the robot used the camera to observe the arrangement and then drew their interpretation of it. The arrangement was changed between robots for each participant to the one captured in the next drawing to be shown. After observing a robot's drawing process, participants received the drawing and rated it as before. Subsequently, they were asked to answer questions about their perception of the robot (Producer) they had just observed drawing (Section 3.3). This procedure was then repeated for the second robot.

Expertise Questions and Debriefing. Afterwards, participants answered questions about their demographic information, art interest, familiarity with AI and familiarity with robotics, using the questionnaires introduced in Section 3.3. To potentially inform the discussion of our findings, they were then briefly interviewed about their overall experiences during the study, and their personal definitions of creativity. They were also encouraged to share any other comments on the

Table 1. Assessed Creativity Means and SDs for Different Levels of PE and Embodiment

PE	Embodiment		
	AxiDraw (SD)	xArm (SD)	Overall Mean (SD)
Product	49.47 (19.69)	47.12 (17.54)	48.29 (18.60)
Product+Process	53.72 (23.41)	51.85 (20.45)	52.78 (21.91)
Product+Process+Producer	59.90 (20.78)	57.15 (18.91)	58.53 (19.83)
Overall Mean	54.36 (21.66)	52.04 (19.34)	53.2 (20.54)

Table 2. Means and SDs of Perceived Robot Attributes Rated on Godspeed Questionnaire Subscales, and the Results of Paired t-Tests Comparing These Attributes for Both Robot Embodiments

Godspeed scale	Embodiment					
	AxiDraw (SD)	xArm (SD)	df	t-Value	p-Value	Cohen's D
Animacy	15.8 (5.33)	18.5 (5.05)	59	−3.74	0.002**	−0.49
Robot Likeability	16.6 (4.31)	18.2 (3.04)	59	−2.92	0.015*	−0.38
Perceived Intelligence	16.7 (3.34)	16.7 (3.26)	59	−0.10	0.920	−0.01
Perceived Safety	11.2 (2.53)	11.6 (2.16)	59	−1.24	0.442	−0.16

**p ≤ 0.01, *p ≤ 0.05, ·p ≤ 0.10. p-values are adjusted using the Holm–Bonferroni method.

experiment. Then we finally revealed that the drawings were originally created by a human artist and that the robots had only reproduced them from a script. We hence offered participants the option to withdraw their consent and data from the study, but no one chose this option.

3.6 Results

The values of our main variable of interest, the assessed Creativity, are reported by PE, by Embodiment, and overall in Table 1. The perceived robot attributes, measured in the Product+Process+Producer level, are reported in Table 2. The participant interest and expertise variables are summarised in Appendix D. We next report the results of our confirmatory analysis in line with our preregistration. This is then complemented by more detailed, exploratory analyses involving participant interest, expertise and perception of robots, amongst others.

3.6.1 Confirmatory Analysis. To analyse the effects of PE and Embodiment on the assessment of Creativity, we conducted a two-way (3 × 2) repeated measures **Analysis of Variance (ANOVA)** using IBM SPSS Statistics (Version 29). The assumptions of ANOVA were adequately met and no corrections for sphericity were required.

The main effect of PE was statistically significant, with a large effect size ($F(2,118) = 13.31$, $p < 0.001$, $\eta_p^2 = 0.18$). The main effect of Embodiment was not statistically significant ($F(1,59) = 3.70$, $p = 0.059$, $\eta_p^2 = 0.06$) and neither was the interaction effect between PE and Embodiment ($F(2,118) = 0.05$, $p = 0.954$, $\eta_p^2 < 0.00$). The effects are visualised in the Figure 6.

We used planned repeated contrasts to investigate the effects of PE further. Results show that, as PE levels increase, Creativity assessments are also higher. In the Product+Process condition, the mean perceived Creativity was 4.49 points higher than in the Product condition and the difference was statistically significant with medium effect size ($F(1,59) = 4.87$, $p = 0.031$, $\eta_p^2 = 0.08$). In the Product+Process+Producer condition, the perceived creativity was 5.75 points higher than in the

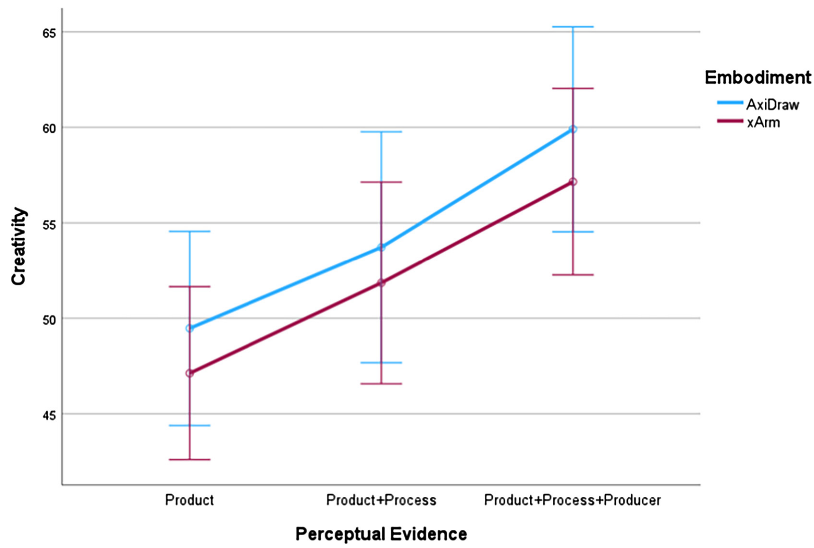


Fig. 6. Interaction plot visualising the relationship of Embodiment and PE on assessed Creativity.

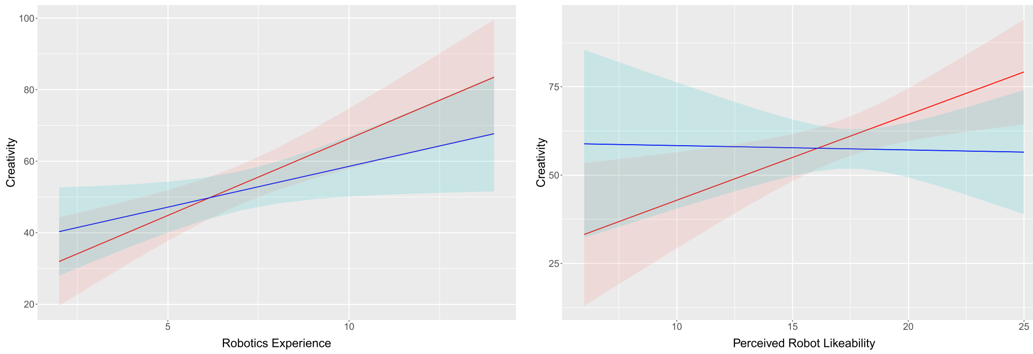
Product+Process condition and the difference was statistically significant with large effect size (F(1,59) = 7.92, p = 0.014, $\eta_p^2 = 0.12$). Reported p-values were adjusted using the Holm–Bonferroni method [1].

3.6.2 Exploratory Analysis. To investigate the relationship of our other variables (interest and expertise variables, perception of robots) and perceived Creativity, we constructed linear mixed models using R Studio [97, 99] and the `lme4` package [8]. Besides fixed effects (population-level relationships), linear mixed models can also account for random effects (individual-level variations), which in some cases allows for more accurate modelling of individual variability compared to traditional analyses of variance. Assumptions were adequately met for all models and all full results and prompts can be found in the OSF Supplementary. Degrees of freedom were approximated using the `lmerTest` package [71] and Satterthwaite's method. All models use the same random structure: separate random intercepts for participants and items (drawings).

We first investigated the effects of our interest and expertise variables (Art Interest, Art Practice, AI Technology Knowledge, AI Usage Experience and Robotics Experience) on perceived Creativity. For fixed effects, the model included PE, Embodiment, their interaction and the five interest and expertise variables. Statistically significant main effects were found for variables AI Usage Experience ($B = -2.24$, 95% CI = [−4.00, −0.49], p = 0.019) and Robotics Experience ($B = 2.61$, 95% CI = [1.05, 4.18], p = 0.003). This indicates that higher AI Usage Experience was connected to lower overall assessments of Creativity, and higher Robotics Experience indicated higher overall assessments of Creativity, while holding other variables constant.

To further investigate these effects separately for each level of PE and Embodiment, we created three additional linear mixed models, one for each level of PE. The last model was also used to inquire into the effects of the perceived robot attributes, which were only recorded on this third level. All of the full models can be found in the OSF Supplementary.

Product. The model included fixed effects for the Embodiment, the five interest and expertise variables, and their interactions with the Embodiment. For the interest and expertise variables and their interactions, the main effect of AI Usage Experience was borderline significant ($B = -2.25$,



(a) PE: Product+Process (b) PE: Product+Process+Producer

Fig. 7. Interaction of (a) Robotics Experience and (b) Perceived Robot Likeability and Embodiment on the assessments of Creativity. AxiDraw as red, xArm as blue, including 95% CI in the shaded areas.

Table 3. PE Level Product+Process+Producer: Significant and Borderline ($p < 0.10$) Fixed Effects and Their Interactions from the Linear Mixed Model

	Estimate [95%CI]	df	p-value
(Intercept)	25.95 [9.27, 15.74]	97.44	0.120
Robot Likeability	2.42 [0.86, 4.04]	83.17	0.007^{**}
Perceived Robot Intelligence	1.47 [−0.04, 4.04]	86.36	$0.079^{\cdot}$
AI Usage Experience	−3.57 [−5.62, −1.47]	82.61	0.002^{**}
Robotics Experience	1.97 [0.15, 3.77]	79.41	0.050^{*}
Embodiment:Robot Likeability	−2.54 [−4.89, −0.19]	63.87	$0.058^{\cdot}$

$^{**}p \leq 0.01$, $^{*}p \leq 0.05$, $^{\cdot}p \leq 0.10$.

95% CI = [−4.40, −0.10], p = 0.051) and the main effect of Robotics Experience was statistically significant ($B = 2.46$, 95% CI = [0.53, 4.36], p = 0.018). These effects are similar to the previous model which did not separate between levels of PE.

Product+Process. The model included the same fixed effects as before. Similarly to the previous level of PE, statistically significant main effects were observed for AI Usage Experience ($B = -2.88$, 95% CI = [−5.21, −0.55], p = 0.023) and Robotics Experience ($B = 4.30$, 95% CI = [2.22, 6.38], $p < 0.001$). Additionally, the interaction effect of Embodiment and Robotics Experience was statistically significant ($B = -2.01$, 95% CI = [−3.53, −0.49], p = 0.015). This interaction indicates that, when perceiving the Product and Process, the effect of Robotics Experience on Creativity assessments was lower when participants assessed the xArm robot Embodiment (Figure 7).

Product+Process+Producer. In addition to the fixed effects considered previously, this model also included fixed effects for our four Godspeed sub-scales (Animacy, Likeability, Perceived Intelligence and Perceived Safety) and their interaction with Embodiment. All statistically significant effects, including borderline significant cases ($p < 0.10$), are reported in Table 3. The results indicate that higher scores on the Godspeed sub-scales Likeability and Perceived Intelligence correspond to higher assessments of perceived Creativity. Likeability also had a borderline significant (p = 0.058) interaction effect with Embodiment (Figure 7), indicating that, under this level of PE, higher scores on robot Likeability were only connected with higher ratings in Creativity when assessed for the Axidraw embodiment. There was no indication of the interaction effect of Embodiment and Robotics Experience ($p = 0.728$) that was present in the previous level of PE, Product+Process.

Differences in Participants' Perception of the Robots. Participants' perception of the Godspeed properties for the two robots are summarised in Table 2. We compared the perception of these robot properties between Embodiments with paired t-tests (Table 2), adjusting the reported p-values with the Holm-Bonferroni method [1]. While differences in Perceived Intelligence and Perceived Safety are almost nonexistent, we found that our participants attributed statistically significantly more Animacy and Likeability to the xArm embodiment, compared to AxiDraw.

Relationships of Creativity questionnaire items. To explore the relationship between the other items (Appendix A) in our creativity questionnaire and perceived Creativity, we calculated correlation matrices, using Pearson correlation coefficient, for each level of PE (see OSF Supplementary for full matrices and 95% CIs). Correlations for the first two levels are similar. All items are fairly correlated with each other and, similar to our pre-study (Appendix B.5), the highest correlations are between the pairs of Creativity and Originality ($r = 0.74-0.80$), Liking and Aesthetic Appeal ($r = 0.84-0.87$), Aesthetic Appeal and Technical Artistic Ability ($r = 0.78-0.82$), and Technical Artistic Ability and Detail ($r = 0.77-0.79$). We note more differences between these two and the third level of PE, especially w.r.t. correlations to Creativity. When comparing correlations while observing Product+Process+Producer to just the Product, participants' Creativity ratings are statistically significantly more correlated with Aesthetic Appeal ($r = 0.79$ vs. $r = 0.64$, $z = 2.63$, $p = 0.001$), and non-significantly more correlated with Technical Artistic Ability ($r = 0.75$ vs. $r = 0.65$, $z = 1.81$, $p = 0.070$), Liking ($r = 0.69$ vs. $r = 0.61$, $z = 1.12$, $p = 0.265$) and Detail ($r = 0.63$ vs. $r = 0.57$, $z = 0.79$, $p = 0.431$). The z-scores and p-values are calculated using the `cocor` package [35] and Steiger's method [111].

4 Discussion

The aim of our study was to investigate whether and how PE and Embodiment affect people's perception of AI Creativity. Our results support our first hypothesis, namely that Creativity assessments would increase as PE accumulates (H1, Section 1). Compared to evaluations based solely on the Product, our participants assessed Creativity higher when also being exposed to the Process, and even higher when the Producer was added, with high statistical significance. Thus, the more was revealed about the very same creative act, the higher our participants assessed its creativity. Contrary to our expectations, our findings do not support our second hypothesis (H2, Section 1); the two robot Embodiments did not result in statistically significant differences in Creativity assessments, neither overall nor within any specific level of PE. We observed small, but not statistically significant, differences in Creativity assessments between Embodiments. These were consistent across all levels of PE, suggesting that they should not be attributed to participants' observations of the robotic Embodiments. In the following, we discuss our findings more closely and compare them to existing empirical work. Due to differences in the measures used, direct comparisons are not possible. For example, while Chamberlain et al. [19] emphasised aesthetic value, we focused on creativity. However, existing work [60, 73, 102] supports that aesthetic value features in the assessment of (AI) creativity especially in the artistic domain. Thus, even indirect comparisons can further our common research questions.

Perception of AI Creativity. In both our studies, the vast majority of participants considered the drawings, their process and their producers creative, as supported by their ratings and the post-study interviews. This suggests that the human perception of creativity is not reserved to other humans, their processes and their artefacts. In contrast, Chamberlain et al. [19] found through a between-subjects design that participants rated the aesthetic value of robot-produced products lower than human-produced, and did not attribute creativity to the robots in terms of the artistic process, nor view them as authors of the drawings. A potential explanation for these differing findings might be a shift in the public perception of creative AI since 2015, when their study was conducted, partly due to development and adoption of increasingly powerful generative AI techniques such as text-to-image diffusion models [117].

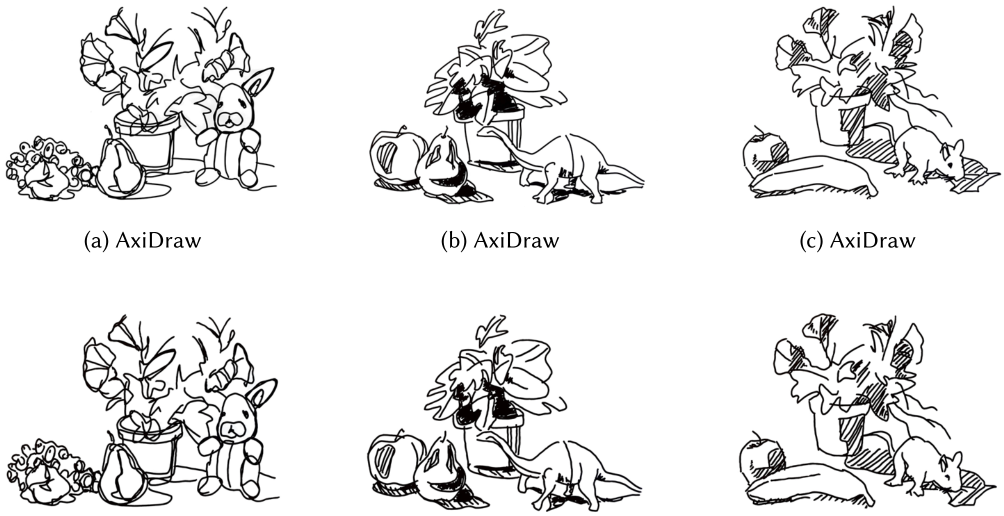
Since we adopted the CAT [4] framework of Creativity assessments, participants were free to use their own subjective definitions of Creativity. From the post-study interviews, it emerged that most associated Creativity with personal expression in drawing, particularly through highlighting, adding, omitting, or abstracting elements from the original subject arrangement. However, a few individuals expressed the view that portrayal art, such as our still life drawings, is inherently uncreative, seeing it as mere copying rather than genuine artistic expression.

Effect of PE. Our findings highlight PE as a significant factor in people's Creativity assessments. This resonates with Chamberlain et al. [19], who observed a significant increase in aesthetic value between the "source information" and "interactive" conditions, paralleling our Product and Product+Process+Producer levels of PE. In addition, we discerned that only exposing the Process in addition to the Product, but not the Producer, significantly increased perceived Creativity. Chamberlain et al. [19] did not distinguish the presence of embodied agents from people's interaction with them, and emphasised this interaction's role in mitigating bias against AI art. Our results in contrast suggest that simply exposing viewers to the embodied agent, even without interaction with them on a joint work, can elevate their perception of creativity. Our interviews also supported this, with some participants noting that just witnessing more of the creative act made them feel more personally involved, leading to more favourable assessments. Additionally, when exposed solely to the Product, some found it challenging to believe that an autonomous robot was the creator.

Our findings are in contrast with Moruzzi [87], who found that an embodied system in the form a robot hand was perceived as less creative compared to an software, though not to a statistically significant degree. However, in their study people only imagined hypothetical scenarios of a software or a robot hand creating art. Our results suggest that witnessing the physical embodied systems in action elicits more pronounced positive reactions leading to higher assessments, compared to merely imagining them.

Effect of Robot Embodiments. Contrary to our expectations, Creativity assessments remained similar across all PE levels, with AxiDraw provoking consistently higher but not statistically significant assessments than xArm (Figure 6). These systematic differences, if not attributed to random chance, suggest that differences in Embodiment subtly influenced the Process and Products. Being well aware of embodiment's potential in shaping creativity, we worked conscientiously to prevent this prior to the experiments. We compared the products of the two robots and revised the drawing software parameters and physical components, such as the penholder, for maximum similarity. Moreover, we designed the drawing script so that both robots draw all lines in the same order, i.e., with Process similarity in mind. A selection of these evaluations is reported in Appendix C, and Figure 8 compares drawings side-by-side. Achieving completely identical products proved challenging, due to practical limitations especially with the xArm, which is not specifically designed for drawing tasks like AxiDraw. Despite these challenges, including xArm was pivotal to compare AxiDraw's simplistic mechanical design with a more organismoid embodiment (cf. Section 3.2).

Our post-study interviews provided qualitative insights into the potential factors explaining these systematic differences. Rather than a clear tendency, we find a strong variety in individual preferences. For instance, participants expressed particularly mixed reactions with respect to AxiDraw's drawing sounds. Some found the distinctive high-pitched mechanical sounds pleasant and almost melodic, likening it to music. Conversely, others found them "irritating", "discomforting"



(a) AxiDraw (b) AxiDraw (c) AxiDraw

(d) xArm (e) xArm (f) xArm

Fig. 8. Comparison of randomly selected example drawings made by AxiDraw (top) and xArm.

or “unsettling”. AxiDraw was also sometimes described as “bare”, “boring”, “ugly” and “primitive” in its appearance. However, few people also preferred its simple, mechanical look. AxiDraw’s drawing technique was typically described as smooth and competent. In contrast, xArm was usually described as more visually impressive and aesthetically pleasing, with its “clean”, “modern” or “futuristic” appearance. However, its drawing process received a more mixed reception compared to AxiDraw. The grip of the penholder was occasionally described as shaky, and sometimes lines were drawn with enough pressure to produce an unpleasant screeching sound on the paper. However, some participants enjoyed xArm’s drawing process, describing its style as more clumsy and giving it more personality. xArm also had more noticeable pauses between drawing commands, which some participants interpreted as the system thinking. While we did not observe direct effects on perceived Creativity for different physical Embodiment, we still believe that embodiment can systematically influence such perception. The qualitative observations here highlight a variety of candidate system characteristics which could play a role. Future research should moreover delve deeper into the individual differences that could categorically affect these perceptions.

During the Product+Process+Producer condition, participants additionally provided quantitative data on their perception of the robot attributes sub-items. Notably, xArm was rated higher than AxiDraw in perceived Animacy and Likeability (Table 2). This supports our initial assumptions in selecting the robots for differences in their organismoid nature (Section 3.2), and confirms that they have been perceived differently. However, these differences did not directly translate into varying creativity assessments; while Likeability was associated with overall higher Creativity assessments only for AxiDraw (Figure 7), it was not sufficient to cause statistically significant differences in the overall assessments between the two Embodiments. It also does not explain the differences in the other PE levels.

We also observed a borderline significant ($p = 0.079$) relationship between Perceived Intelligence and Creativity, indicating that higher perception of intelligence was associated with higher assessments of Creativity. In our findings, participants perceived both robots equally intelligent (see Table 2), rating them both slightly above the scale’s midpoint (16.7 vs. 15). This suggests that, while participants deemed the robots somewhat intelligent, their perception of intelligence could be more attributed to the drawing task itself, rather than to specific robot Embodiment.

While Chamberlain et al. [19] discovered that anthropomorphism influenced overall assessments of aesthetic value, we did not observe similar effect for Animacy, even though the two concepts are arguably intertwined for our robots. However, this might be due to our robots not being sufficiently human-like in their appearance and behaviour to elicit a similar response. Although xArm was perceived as having higher Animacy compared to AxiDraw, it lacked several characteristics from the robots studied by Chamberlain et al. [19]. Most notably, their robots were deliberately programmed by the artist Patrick Tresset to perform numerous theatrical actions and appear more alive and human-like. Despite mixed results, impressions of anthropomorphism and animacy could partly account for the aesthetic and creative value attributed to machine-made art, potentially through eliciting emphatic responses, as discussed in related work (Section 2). Future studies should explore these effects by including machines that not only look more humanoid, but also behave in a more human-like manner.

Interest and Expertise. We did not observe Art Interest influencing Creativity assessments under any condition. Previously, art interest has been found to influence people's aesthetic assessments when evaluating human art [74, 77] and, notably, Chamberlain et al. [19] reproduced this effect for machine-made art. The discrepancy in our findings compared to Chamberlain et al. [19] could arise from differences in measures. We employed the validated Art Interest sub-scale of VAIAK [110], which measures a general interest in art, including statements like "I enjoy talking about art with others" and "I am interested in art". In contrast, Chamberlain et al. [19] focused on two questions about specific behaviours of going to art galleries or owning art books. While similar questions are also included in the VAIAK sub-scale, the latter might capture art expertise more broadly. Future studies should revisit these effects of art interest, paired with validated measures of art education and experience.

In contrast to Art Interest, AI Technology Knowledge and Robotics Experience showed significant effects on the Creativity assessment ($p \leq .051$) both overall and across all PE levels. Interestingly, their directions of influence were opposed. On the one hand, increased AI Usage Experience corresponded to lower Creativity assessments, possibly suggesting that our simplistic croquis drawings did not meet the expectations of those presumably familiar with advanced generative AI models. On the other hand, more Robotics Experience was associated with higher Creativity assessments. Since this effect persisted even in the Product conditions, it cannot be solely attributed to the observation of the physical robots. We hypothesise that individuals familiar with robotics are more impressed by robots in novel situations, as many participants expressed that drawing robots were a very novel concept to them. We also observed an interaction effect between Robotics Experience and Embodiment. Its influence on Creativity assessments was larger for the AxiDraw robot, but only on the PE level Product+Process. These findings are important avenues for future replication and research, especially with more diverse robots and art forms.

Deception. In our studies, we employed deception, leading participants to believe that the predetermined drawings were innovated by autonomous robots, although AI was not actually used. Using deception was crucial to avoid potential confounding factors related to participants beliefs about how the drawings were created. If we had been more vague and only informed participants that the robots were generating the drawings without explicitly mentioning AI, participants might have formed varying assumptions about the exact origins of the drawings. Some might have assumed that AI was involved, while others might have thought that humans were behind the original process, and they could potentially even change their assumptions between conditions. This ambiguity could introduce confounders, such as biases against machines or people attributing creativity to the imagined human artist behind the drawings, instead of the robot.

While we did not formally assess each participant's belief in the deception, their reactions during the studies provided clear evidence of its effectiveness. The overwhelming majority expressed surprise upon the reveal of the deception, with only two participants expressing vague suspicions about the "AI system". This was further supported by some participants speculating during the study about the mechanisms behind the robots' drawing capabilities. Various speculations were offered: some hypothesised the use of a "simple rule based AI", others surmised that the system might lack object recognition, functioning merely by following grayscale gradients. Additionally, there were conjectures about the extent of datasets and specific techniques about identifying features such as colours or shadows. These speculations clearly indicate that the participants were under the impression that some kind of an AI was present. The perceived Intelligence of both robots was also assessed above the midpoint of the used Godspeed scale (16.7 vs. 15), further supporting the belief in an AI. These points strongly suggest that our findings are applicable to real AI contexts.

5 Study Limitations

Our study has particular limitations that could provide valuable insights and leads for future work.

Generalisability. Our stimuli consisted of simple, croquis-style still life drawings. While croquis is a well-established artistic practice, it is quick and sketchy by definition, and requires concentrating detail on a subject's most important aspects. Some noted that this simplicity made creativity assessments more challenging. To investigate whether our findings generalise, future studies could include work with greater precision and a more homogeneous distribution of detail. Of course, the challenge here is to balance this with robots' affordances and drawing times. Echoing Chamberlain et al. [19], we also encourage future related studies to investigate abstract art as stimuli. Abstract art requires different sensorimotor affordances which, when observed, could affect the assessment of creativity. Abstract artworks have also been found to reduce the negative bias humans generally have towards AI-art [22].

We conducted our study in a lab setting affording maximum control for this early-stage experiment. However, for increased ecological validity, we support future studies to be conducted both in the lab and in the wild. Museum spaces, as explored, e.g., by Gemeinboeck and Saunders [42] as well as Chamberlain et al. [19], may offer a good trade-off between controllability and ecological validity, but may introduce participant selection bias.

Autonomy in Subject Selection. Some participants expressed that a significant portion of the creativity in still life drawing stems from the selection and arrangement of the included objects. This is supported by both theoretical and empirical work, stressing the importance of autonomy for the attribution of creativity to machines [29, 60, 83]. In our study, this arrangement was predetermined for the robots. We consider this unproblematic as (i) we were interested in relative differences in creativity assessment and (ii) encouraged relative assessments in our study introductions.

Association With Drawing Style. In order to prevent mere exposure effects, we deliberately asked our commissioned artist to produce drawings in different styles. This however may have introduced unintended effects as some participants, especially while observing the robots, seemed to associate distinct drawing styles with specific robots. Such associations could introduce bias if participants favoured one style over another. While we believe that our counterbalancing strategy of drawings and embodiments should minimise these effects, we cannot control for individual differences in style preference. Future studies employing a similar design might benefit from including multiple drawings with different styles for each embodiment across all PE levels. This could help to demonstrate that styles are not specific to individual robots.

Sequential Exposure. Having all participants encounter the PE levels in the same sequential order was necessary for us to avoid leaking impressions from richer into lower levels, potentially biasing people's assessments (see also Section 3.1). However, this procedure has the potential to introduce systematic noise and order effects. As participants progress through the session, and consequently the PE levels, they become more familiar with the drawings and the assessment process. This could make them more adept at e.g. identifying details, lead them into forming preferences for certain details, or make them give higher ratings due to being more familiar with the drawings. While these could in principle affect Creativity assessments in later conditions, increased awareness of artistic details is more likely cancelled out by the additional discovery of small flaws, or reduced perception of creativity due to reduced novelty.

Overall, we hypothesised that a fixed order in presenting the PE levels would introduce less bias compared to a counterbalanced order. An alternative approach could have been a between-subjects design with three groups. However, such design would require significantly larger sample size to achieve comparable statistical power. Additionally, a between-subjects approach might introduce biases due to individual differences. While a matched-group design could potentially address this, it is impractical in our context where we are only initially exploring which individual differences might influence perception of Creativity. Nevertheless, we advocate for future research to consider different study designs, including between-subjects and counterbalancing the order of PE levels, to investigate the impact of exposure order on participant perceptions.

Exposure Time. While evaluating only the Product, participants spend considerably less time observing the stimuli when compared to additionally watching videos or physical drawing processes in the other PE levels. Although the impact of exposure time on creativity is uncertain, we deem it less problematic in our study given the croquis-like nature of our stimuli, which does not reveal considerably more detail upon extended viewing. For future work embracing different stimuli however, we recommend controlling for exposure time.

Choice of Scales. We did not explicitly measure participants' art knowledge, since the Art Interest sub-scale of VAIAK has been shown to correlate highly with general art knowledge [110]. Crucially though, art education and knowledge have been shown to influence people's aesthetic assessments [19, 74]. Going forward, it might thus be beneficial to employ purposeful sampling and additionally investigate our research questions on creativity on individuals with strong art education and knowledge.

Similarly, we did not explicitly assess the perceived anthropomorphism of the robots, though we expect it would correlate strongly with the Animacy sub-scale, which we incorporated. However, we suggest to investigate perceived anthropomorphism in the future, especially when including humanoid robots that more closely mimic human-like shapes, movements and behaviours.

We used broad and general scales to assess participants' familiarity and experience with AI and robotics. This approach was partly due to the absence of validated instruments at the time of our study and partly a choice to use an exploratory approach to gather insights for formulating hypotheses for confirmatory research in future. Nevertheless, this approach limited our understanding of the exact nature of individual differences in the current study. For instance, it remains unclear if participants' reported AI usage included prior use of creative AI or creativity support systems. However, during the interviews, participants did not report any such usage when queried about factors influencing their ratings of creativity. Future research should adopt more comprehensive and systematic qualitative measures for a deeper understanding of these individual differences.

Robot Limitations. Some of our limitations arose from technical difficulties, particularly with the xArm robot. As discussed in Section 4, replicating the high drawing precision of AxiDraw with xArm proved challenging. Moreover, xArm occasionally malfunctioned while operating, the worst cases leading to the invalidation and subsequent replacement of two sessions. Because of its custom software and our addition of physical components, the robot also required additional calibration between different drawings to ensure consistent experiences for all participants. Both of these issues may have contributed to the systematic, but non-significant, differences in Creativity assessments. While most of these issues are commonplace in HRI studies, we still note that ambitions in experiment design must be balanced with the technical affordances of the robots at hand, and challenges of using robots not originally designed for drawing tasks should not be underestimated.

Decoupling Embodiment Effects. One core characteristic of our study design, adapted from Linkola et al. [78], is that we tightly controlled the creative process throughout the conditions. Given the immense role of embodiment for cognition [14, 15, 94], one could criticise that holding the process constant introduces a tension between the affordances of the system, as communicated through its embodiment, and the perceived action. Put differently, a change in embodiment is not perceived as a change in the creative process. This effect should be further explored in future work, potentially lending more value to alternative experiment designs akin to those employed by Chamberlain et al. [19].

6 Implications of our Findings

The findings and limitations of this study have a wide range of implications for the design of creative AI performances, shaping public awareness of biases in the perception of creative AI, and contributing to scientific progress on the Lovelace and embodiment effects across disciplines, in terms of reflecting on past and shaping future research. They affect and provide value for researchers across Computational Creativity, HCI, HRI and Psychology.

Supporting the effect of PE on people's assessment of creativity (*H1*, Section 1), we provide empirical support for established design patterns in "computational creativity theatre" [27], i.e., the deliberate staging of a creative AI system to heighten observers' perception of its creativity. Our findings particularly support previously held beliefs in not only showcasing the final product, but also equipping a system with the facilities to enact the creative process, and embodying it physically or virtually. Crucially, while existing work proposes to increase the perception of creativity by framing the process and producer with various kinds of detailed information [30], we show that mere exposure to the process and producer per se already has a positive effect. While theoretical work promotes the choice of humanoid over different embodiment morphologies [27, 28], our findings on *H2* (Section 1) do not support this. However, considering the strong support from HRI studies, we recommend further inquiry into morphology as a moderator of embodiment effects in the attribution of creativity, supported by our reflection on results (Section 4) and study limitations (Section 5).

While these insights can support the design and staging of systems which are more likely perceived as creative, we hope that they equally shape public awareness of our biases when confronted with creative AI, empowering people in reflecting on their experiences. This is the more important now that creative AI takes a more prominent role in our everyday lives, awakes hopes and fears [67, 117], features in gallery exhibition and auction houses [11], and is used by corporations to manifest their position at the forefront of AI research and its commodification. We hope these insights to become commonplace just like the fact that presenting an artwork in an exhibition space positively influences our evaluation [92], an effect that has recently been revisited in the context of creative AI [11, 91]. Echoing Chamberlain et al. [19], we believe that such research on the Lovelace effect is critical for understanding how creative AI can and will be assimilated into our society.

From a scientific perspective, our findings promote a more critical view on past and future work investigating the Lovelace effect. Most importantly, they teach us that, when comparing how different systems affect people's creative assessments, participants' PE about the creative act must be controlled for. Vice versa, we should apply scrutiny to studies which fail to do so, but in which conditions also differ in the level of PE they afford. In such situations, PE may act as a moderator between the factor under investigation and the creativity assessment. This is supported with respect to other outcomes by theoretical and empirical work in HRI (e.g., [55, 70]).

Our findings also allow us to differentiate existing findings and thus raise new research questions. Our comparison to Chamberlain et al's [19] results (Section 4) for instance provokes the question whether the observation of the AI as embodied agent, the interaction with them, or both, mitigate human bias against creative AI. Our results even invite to question whether comparisons of human and AI creativity relying only on the product, or on the product and process, are reasonable, as we expect people to fill in the missing evidence through their imagination, likely introducing the more noise the less evidence is offered. Our findings on the effect of Robotics Experience and AI Usage Experience highlight and isolate specific factors that moderate disparities in people's assessment of creativity in AI. They could also imply that, as the general public becomes more educated and experienced in AI or robotics, their assessments of AI creativity might shift. This provides two distinct motivations for further studies providing deeper insights into the effect of experience and expertise on the perception of creativity in AI. While our findings do not support an embodiment effect on people's assessment of machine creativity, we encourage continued research into this subject since such effects would be extremely informative of system design and selection in creative practice. To this end, the discussion and interpretation of our findings (Section 4) provide valuable starting points for future hypotheses.

Finally, our study limitations can serve as recommendation for such future work. The issues faced with the xArm robot highlight the challenges of integrating physical technology with this line of research. They can inform future studies to include a control or pre-study to prevent the embodiment from influencing product or process conditions, even if difference seem negligible and well controlled for. Moreover, they highlight the constraints which the features and cost of available hardware put on the types of experiments that we can conduct; for instance, engineering and instructing a full humanoid robot to draw at the same precision as the AxiDraw plotter is extremely challenging. This puts hard limitations on how much of the creative process can be controlled for, but also highlights opportunities in collaborations with expert roboticists, both from research and arts.

Our study focuses on the observation of (allegedly) autonomously creative robots, rather than a richer interaction with co-creative systems. This was a deliberate decision to avoid introducing the interaction task as a well-acknowledged additional moderator [55]. With the effect of PE supported, we consider this study an important stepping stone toward future research on the Lovelace effect in human–machine co-creativity specifically. The attribution of creativity is of core relevance to co-creativity research, as co-creative systems are distinguished from mere creativity support by virtue of exhibiting some degree of creative autonomy [64], and it is widely assumed that engagement with a system is predicted by perceived creativity. Even without embracing an interactive task, the present study already holds implications for at least two important scenarios in co-creative interaction: the first engagement with the system, and ongoing engagement in alternating co-creativity. Especially when encountering creative systems that invite but do not necessitate co-creation, e.g. in music [104, 106] or art performance [105], the potential partner can assess the creativity of the system in action before deciding to engage with it. Moreover, in alternating co-creativity [65], individual interaction steps can comprise considerable amounts of independent creative work, and offer an opportunity for the human partner to re-assess the creativity of the system as condition for their continuous engagement in the next alternation. We moreover hold that the present study provokes important questions for future co-creativity research. To only name one, popular user interface design patters for co-creative systems rely on sharing only the product with the interaction partner. While this could be considered as steps in a macro-level process, our findings raise the question whether sharing the creation process more fluidly, or even interacting with the embodied system in the same space, could support the perception of creativity and consequently acceptance of and engagement with such systems as creative partners.

7 Conclusion and Future Work

In this article, we present an empirical study of the Lovelace effect, i.e., people's attribution of creativity to artificial systems. More specifically, we investigate the effect of two variables: PE, which aspects of an AI system are perceived, and the system's embodiment. Informed by the well-established four perspectives on creativity [58, 85, 100], PE has been distinguished into the product, process and producer. Following calls from Computational Creativity and HRI (e.g., [55, 78]) embodiment as characteristic of the producer has been distinguished in terms of morphology, corresponding to a mechanistic plotter and organismoid robot arm. The influence of these variables was investigated in a 3×2 complete factorial experiment on artistic creativity (drawings) under controlled lab conditions. People's perception of creativity was assessed by adopting the well established CAT for this particular scenario. Our findings reveal a significant effect of PE, with additional evidence of the process and producer positively affecting people's perception of creativity. In other words, the human assessment of machine creativity seems to increase as more is revealed about the creation process and the producer. At the same time, the influence of embodiment as characteristic of the producer remains ambiguous. Additionally, we identified several other factors in participants' experience with respect to AI usage and robotics, as well as robot attitudes (Likeability and Perceived Intelligence), which may play a role in assessing these systems and their products. Amongst other implications, our results support existing patterns for the design and staging of systems to be more likely perceived as creative, and clearly highlight the necessity to control for PE in future evaluations of creative AI systems, for instance, when aiming to determine the most suitable systems to be accepted by people in creative partnerships.

Future work can learn from our limitations and gain inspiration from our findings. We particularly advocate the complement quantitative inquiry with more exploratory, qualitative work to add granularity to our understanding of the factors governing our assessment of creativity in embodied AI systems. Especially the individual differences noted by our exploratory analyses warrant further examination through more detailed and in-depth qualitative inquiries. Moreover, future work should probe the generalisability of our findings. To this end we should investigate whether the same effects hold for other artistic styles or genres, such as abstract art or music. Similarly, whether the findings generalise to other creative agents besides robots, such as people or animals, should be investigated. Other study designs should also be explored. Additional research should also be dedicated to identifying moderators for these effects, e.g., order and time of exposure to PE. Different, more interactive tasks should be employed, thus introducing interaction scenarios as additional moderator as highlighted by Hoffmann et al. [55]. We also encourage researchers to replicate our findings with actual AI systems and no deception. For all these endeavours, we advocate closer collaborations between researchers and artists who specialise in building highly controllable creative machines.

At these times when creative AI is having an unprecedented impact on our everyday lives, we consider such studies of paramount importance to better understand our perception of AI creativity as a determinant of our future interaction.

Author Contributions

We highlight individual contributions using the Contributor Roles Taxonomy [31], numbers representing author positions: Conceptualisation (1, 2, 3, 6), Methodology (1, 2, 3, 6), Software (2, 4), Investigation (1), Data Curation (1), Formal Analysis (1, 6), Visualisation (1), Writing (1, 2, 4, 5, 6), Resources (5, 6), Funding Acquisition (2, 3, 5 ,6) and Supervision (6).

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgements

We thank artist Mikaela Hongell for her work on the original drawings, our participants for their interest in our studies, and our reviewers for their kind words and constructive suggestions. It moreover benefited from infrastructure provided by Aalto University's Radical Creativity initiative.

References

[1] Hervé Abdi. 2010. Holm's sequential Bonferroni procedure. *Encyclopedia of Research Design* 1, 8 (2010), 1–8.

[2] Kat Agres, Jamie Forth, and Geraint A. Wiggins. 2016. Evaluation of musical creativity and musical metacreation systems. *Computers in Entertainment (CIE)* 14, 3 (2016), 1–33.

[3] Teresa M. Amabile. 1979. Effects of external evaluation on artistic creativity. *Journal of Personality and Social Psychology* 37, 2 (1979), 221.

[4] Teresa M. Amabile. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology* 43, 5 (1982), 997–1013.

[5] Teresa M. Amabile. 2018. *Creativity in Context: Update to the Social Psychology of Creativity*. Routledge.

[6] Katrin Auspurg and Thomas Hinz. 2014. *Factorial Survey Experiments*, Vol. 175. Sage Publications.

[7] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1 (2009), 71–81.

[8] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

[9] Debarun Bhattacharjya. 2016. Preference models for creative artifacts and systems. In *Proceedings of the International Conference on Computational Creativity (ICCC)*. Citeseer, 52–59.

[10] Margaret A. Boden. 2015. Foreword: How computational creativity began. In *Computational Creativity Research: Towards Creative Machines*. Tarek R. Besold, Marco Schorlemmer, and Alan Smaill (Eds.), Atlantis Press, 5–13.

[11] Ian Bogost. 2019. The AI-Art Gold Rush is Here. Retrieved from https://www.theat-lantic.com/technology/archive/2019/03/ai-createdart-invades-chelsea-gallery-scene/584134/

[12] Paolo Bory. 2019. Deep new: The shifting narratives of artificial intelligence from Deep Blue to AlphaGo. *Convergence* 25, 4 (2019), 627–642.

[13] Mason Bretan and Gil Weinberg. 2016. A survey of robotic musicianship. *Communications of the ACM* 59, 5 (2016), 100–109.

[14] Rodney A. Brooks. 1990. Elephants don't play chess. *Robotics and Autonomous Systems* 6, 1-2 (1990), 3–15.

[15] Rodney A. Brooks. 2018. Intelligence without reason. In *The Artificial Life Route to Artificial Intelligence*. Routledge, 25–81.

[16] Astrid Carolus, Martin Koch, Samantha Straka, Marc Erich Latoschik, and Carolin Wienrich. 2023. MAILS–Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change-and meta-competencies. arXiv:2302.09319. Retrieved from https://arxiv.org/abs/2302.09319

[17] Colleen M. Carpinella, Alisa B. Wyman, Michael A. Perez, and Steven J. Stroessner. 2017. The robotic social attributes scale (RoSAS): Development and validation. In *Proceedings of the 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 254–262.

[18] Rebecca Chamberlain, Daniel Berio, Veronika Mayer, Kirren Chana, Frederic Fol Leymarie, and Guido Orgs. 2022. A dot that went for a walk: People prefer lines drawn with human-like kinematics. *British Journal of Psychology* 113, 1 (2022), 105–130.

[19] Rebecca Chamberlain, Caitlin Mullin, Bram Scheerlinck, and Johan Wagemans. 2018. Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychology of Aesthetics, Creativity, and the Arts* 12, 2 (2018), 177.

[20] John William Charnley, Alison Pease, and Simon Colton. 2012. On the notion of framing in computational creativity. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 77–81.

[21] Chuansheng Chen, Joseph Kasof, Amy J. Himsel, Ellen Greenberger, Qi Dong, and Gui Xue. 2002. Creativity in drawings of geometric shapes: A cross-cultural examination with the consensual assessment technique. *Journal of Cross-Cultural Psychology* 33, 2 (2002), 171–187.

[22] Salvatore G. Chiarella, Giulia Torromino, Dionigi M. Gagliardi, Dario Rossi, Fabio Babiloni, and Giulia Cartocci. 2022. Investigating the negative bias towards artificial intelligence: Effects of prior assignment of AI-authorship on the aesthetic appreciation of abstract paintings. *Computers in Human Behavior* 137 (2022), 107406.

[23] Francesca Ciardo, Davide De Tommaso, and Agnieszka Wykowska. 2022. Joint action with artificial agents: Human-likeness in behaviour and morphology affects sensorimotor signaling and social inclusion. *Computers in Human Behavior* 132 (2022), 107237.

[24] Simon Colton. 2008. Creativity versus the perception of creativity in computational systems. In *Proceedings of the AAAI Spring Symposium: Creative Intelligent Systems*, Vol. 8, 7.

[25] Simon Colton, John William Charnley, and Alison Pease. 2011. Computational creativity theory: The FACE and IDEA descriptive models. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 90–95.

[26] Simon Colton, Jakob Halskov, Dan Ventura, Ian Gouldstone, Michael Cook, and Blanca Pérez Ferrer. 2015. The painting fool sees! New projects with the automated painter. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 189–196.

[27] Simon Colton, Jon McCormack, Michael Cook, and Sebastian Berns. 2020. Creativity theatre for demonstrable computational creativity. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 288–291.

[28] Simon Colton, Alison Pease, and Rob Saunders. 2018. Issues of authenticity in autonomously creative systems. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 272–279.

[29] Simon Colton and Geraint A. Wiggins. 2012. Computational creativity: The final frontier? In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, Vol. 12, 21–26.

[30] Michael Cook, Simon Colton, Alison Pease, and Maria Teresa Llano. 2019. Framing in computational creativity-A survey and taxonomy. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 156–163.

[31] CASRAI CRediT. 2022. Contributor roles taxonomy. Retrieved from https://credit.niso.org/

[32] Genevieve M. Cseh and Karl K. Jeffries. 2019. A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychology of Aesthetics, Creativity, and the Arts* 13, 2 (2019), 159.

[33] Theresa Rahel Demmer, Corinna Kühnapfel, Joerg Fingerhut, and Matthew Pelowski. 2023. Does an emotional connection to art really require a human artist? Emotion and intentionality responses to AI-versus human-created art and impact on aesthetic experience. *Computers in Human Behavior* 148 (2023), 107875.

[34] Oliver Deussen, Thomas Lindemeier, Sören Pirk, and Mark Tautzenberger. 2012. Feedback-guided stroke placement for a painting machine. In *Proceedings of the Eighth Annual Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging*.

[35] Birk Diedenhofen and Jochen Musch. 2015. cocor: A comprehensive solution for the statistical comparison of correlations. *PloS One* 10, 4 (2015), e0121945.

[36] Dennis Dutton. 2003. Authenticity in art. In *The Oxford Handbook of Aesthetics*, J. Levinson (Ed.), Oxford University Press, 258–274.

[37] Denis Dutton. 2009. *The Art Instinct: Beauty, Pleasure, & Human Evolution*. Oxford University Press.

[38] Ziv Epstein, Aaron Hertzmann, Laura Herman, Robert Mahari, Morgan R. Frank, Matthew Groh, Hope Schroeder, Amy Smith, Memo Akten, Jessica Fjeld, et al. 2023. Art and the science of generative AI: A deeper dive. arXiv:2306.04141. Retrieved from https://arxiv.org/abs/2306.04141

[39] Ziv Epstein, Sydney Levine, David G. Rand, and Iyad Rahwan. 2020. Who gets credit for AI-generated art? *Iscience* 23, 9 (2020), 101515.

[40] David Freedberg and Vittorio Gallese. 2007. Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences* 11, 5 (2007), 197–203.

[41] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the landscape of creativity support tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 1–18.

[42] Petra Gemeinboeck and Rob Saunders. 2013. Creative machine performance: Computational creativity and robotic art. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 215–219.

[43] Petra Gemeinboeck and Rob Saunders. 2022. Moving beyond the mirror: Relational and performative meaning making in human–robot communication. *AI & Society* 37, 2 (2022), 549–563.

[44] Vlad PetreGlăveanu. 2013. Rewriting the language of creativity: The Five A's framework. *Review of General Psychology* 17, 1 (2013), 69–81.

[45] Michael A. Goodrich and Alan C. Schultz. 2008. Human–robot interaction: A survey. *Foundations and Trends® in Human–Computer Interaction* 1, 3 (2008), 203–275.

[46] Christian Guckelsberger, Anna Kantosalo, Santiago Negrete-Yankelevich, and Tapio Takala. 2021. Embodiment and computational creativity. In *Proceedings of the International Conference on Computational Creativity*.

[47] Christian Guckelsberger, Christoph Salge, and Simon Colton. 2017. Addressing the "why?" in computational creativity: A non-anthropocentric, minimal model of intentional creative agency. In *Proceedings of the International Conference on Computational Creativity (ICCC)*.

[48] Joy Paul Guilford. 1967. *The Nature of Human Intelligence*. McGraw-Hill.

[49] Jörg Marvin Gülzow, Patrick Paetzold, and Oliver Deussen. 2020. Recent developments regarding painting robots for research in automatic painting, artificial creativity, and machine learning. *Applied Sciences* 10, 10 (2020), 3396.

[50] Hatice Gunes, Frank Broz, Chris S. Crawford, Astrid Rosenthal-von der Pütten, Megan Strait, and Laurel Riek. 2022. Reproducibility in human-robot interaction: Furthering the science of HRI. *Current Robotics Reports* 3 (2022), 281–292.

[51] Beth A. Hennessey. 1994. The consensual assessment technique: An examination of the relationship between ratings of product and process creativity. *Creativity Research Journal* 7, 2 (1994), 193–208.

[52] Laura Mariah Herman and Angel Hsing-Chi Hwang. 2022. In the eye of the beholder: A viewer-defined conception of online visual creativity. *New Media & Society* (2022).

[53] Laura Mariah Herman and Caterina Moruzzi. 2022. Exploring embodiment's role in creativity with live artistic performances. In *Proceedings of the Workshop on the Role of Embodiment in the Perception of Human & Human & Artificial Creativity (TREPHAC 2022) Co-located with the 13th International Conference on Computational Creativity (ICCC 2022),* 1–8. CEUR Workshop Proceedings.

[54] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS '20)*, 6840–6851.

[55] Laura Hoffmann, Nikolai Bock, and Astrid M. Rosenthal vd Pütten. 2018. The peculiarities of robot embodiment (EmCorp-scale) development, validation and initial test of the embodiment and corporeality of artificial agents scale. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 370–378.

[56] Joo-Wha Hong, Katrin Fischer, Yul Ha, and Yilei Zeng. 2022. Human, I wrote a song for you: An experiment testing the influence of machines' attributes on the AI-composed music evaluation. *Computers in Human Behavior* 131 (2022), 107239.

[57] Nanna Inie, Jeanette Falk, and Steve Tanimoto. 2023. Designing participatory AI: Creative professionals' worries and expectations about generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–8.

[58] Anna Jordanous. 2012. A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation* 4 (2012), 246–279.

[59] Anna Jordanous. 2016. Four PPPPerspectives on computational creativity in theory and in practice. *Connection Science* 28, 2 (2016), 194–216.

[60] Anna Jordanous and Bill Keller. 2016. Modelling creativity: Identifying key components through a corpus-based approach. *PloS One* 11, 10 (2016), e0162959.

[61] Ankur Joshi, Saket Kale, Satish Chandel, and D. Kumar Pal. 2015. Likert scale: Explored and explained. *British Journal of Applied Science & Technology* 7, 4 (2015), 396.

[62] Jean-Luc Jucker, Justin L. Barrett, and Rafael Wlodarski. 2014. "I just don't get it": Perceived artists' intentions affect art evaluations. *Empirical Studies of the Arts* 32, 2 (2014), 149–182.

[63] Anna Kantosalo and Tapio Takala. 2020. Five C's for human-computer co-creativity-an update on classical creativity perspectives. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 17–24.

[64] Anna Kantosalo, Jukka M. Toivanen, Ping Xiao, and Hannu Toivonen. 2014. From isolation to involvement: Adapting machine creativity software to support human-computer co-creation. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 1–7.

[65] Anna Kantosalo and Hannu Toivonen. 2016. Modes for creative human-computer collaboration: Alternating and task-divided co-creativity. In *Proceedings of the 7th International Conference on Computational Creativity*, 77–84.

[66] James C. Kaufman, John Baer, Jason C. Cole, and Janel D. Sexton. 2008. A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal* 20, 2 (2008), 171–178.

[67] Hyung-Kwon Ko, Gwanmo Park, Hyeon Jeon, Jaemin Jo, Juho Kim, and Jinwook Seo. 2023. Large-scale text-to-image generation models for visual artists' creative works. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 919–933.

[68] Terry K. Koo and Mae Y. Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* 15, 2 (2016), 155–163.

[69] Justin Kruger, Derrick Wirtz, Leaf Van Boven, and T. William Altermatt. 2004. The effort heuristic. *Journal of Experimental Social Psychology* 40, 1 (2004), 91–98.

[70] Laura Kunold, Nikolai Bock, and Astrid Rosenthal-von der Pütten. 2023. Not all robots are evaluated equally: The impact of morphological features on robots' assessment through capability attributions. *ACM Transactions on Human-Robot Interaction* 12, 1 (2023), 1–31.

[71] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82 (2017), 1–26.

[72] Carolyn Lamb, Daniel G. Brown, and Charles Clarke. 2015. Human competence in creativity evaluation. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 102–109.

[73] Carolyn Lamb, Daniel G. Brown, and Charles L. A. Clarke. 2018. Evaluating computational creativity: An interdisciplinary tutorial. *ACM Computing Surveys (CSUR)* 51, 2 (2018), 1–34.

[74] Helmut Leder, Gernot Gerger, Stefan G. Dressler, and Alfred Schabmann. 2012. How art is appreciated. *Psychology of Aesthetics, Creativity, and the Arts* 6, 1 (2012), 2.

[75] Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User experience design professionals' perceptions of generative artificial intelligence. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*.

[76] David Liljequist, Britt Elfving, and Kirsti Skavberg Roaldsen. 2019. Intraclass correlation–A discussion and demonstration of basic features. *PloS One* 14, 7 (2019), e0219854.

[77] Annukka K. Lindell and Julia Mueller. 2011. Can science account for taste? Psychological insights into art appreciation. *Journal of Cognitive Psychology* 23, 4 (2011), 453–475.

[78] Simo Linkola, Christian Guckelsberger, Tomi Männistö, and Anna Kantosalo. 2022. How does embodiment affect the human perception of computational creativity? An experimental study framework. In *The Role of Embodiment in the Perception of Human & Artificial Creativity 2022: Workshop at Proceedings of the International Conference on Computational Creativity*.

[79] Haiying Long and Jue Wang. 2022. Dissecting reliability and validity evidence of subjective creativity assessment: A literature review. *Educational Psychology Review* 34, 3 (2022), 1399–1443.

[80] Todd Lubart, Dario Esposito, Alla Gubenko, and Claude Houssemand. 2021. Creativity in humans, robots, humbots. *Creativity. Theories–Research-Applications* 8, 1 (2021), 23–37.

[81] John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon. 2006. A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine* 27, 4 (2006), 12–14.

[82] Tom McClelland. 2022. Misembodied experiences of artificial art. In *The Role of Embodiment in the Perception of Human & Artificial Creativity 2022: Workshop at Int. Conf. Computational Creativity*.

[83] Jon McCormack, Toby Gifford, and Patrick Hutchings. 2019. Autonomy, authenticity, authorship and intention in computer generated art. In *Proceedings of the International Conference on Computational Intelligence in Music, Sound, Art and Design (part of EvoStar)*. Springer, 35–50.

[84] Luigi Federico Menabrea. 1842. *Sketch of the Analytical Engine Invented by Charles Babbage, With Notes Upon the Memoir by the Translator Ada Augusta, Countess of Lovelace*. Bibliotheque Universelle de Geneve.

[85] Ross L. Mooney. 1963. A conceptual model for integrating four approaches to the identification of creative talent. In *Scientific Creativity: Its Recognition and Development*, 331–340.

[86] Caterina Moruzzi. 2021. Measuring creativity: An account of natural and artificial creativity. *European Journal for Philosophy of Science* 11, 1 (2021), 1.

[87] Caterina Moruzzi. 2022. The (artificial) physicality of creativity: How embodiment influences perceptions of creativity. In *Proceedings of 13th International Conference on Computational Creativity*.

[88] Caterina Moruzzi. 2022. Perceptions of creativity in artistic and scientific processes. In *Proceedings of the Conference on Computation, Communication, Aesthetics & X (xCoAx).*

[89] Caterina Moruzzi and Laura Mariah Herman. 2022. Exploring embodiment's role in creativity with live artistic performances. In *The Role of Embodiment in the Perception of Human & Artificial Creativity 2022: Workshop at the International Conference on Computational Creativity*.

[90] Martin Mumford and Dan Ventura. 2015. The man behind the curtain: Overcoming skepticism about creative computing. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 1–7.

[91] Simone Natale and Leah Henrickson. 2022. The Lovelace effect: Perceptions of creativity in machines. *New Media & Society* 26, 4 (2022), 1909–1926.

[92] Brian O'doherty. 1999. *Inside the White Cube: The Ideology of the Gallery Space*. University of California Press.

[93] Kohei Ogawa, Koichi Taura, and Hiroshi Ishiguro. 2012. Possibilities of androids as poetry-reciting agent. In *Proceedings of the International Symposium on Robot and Human Interactive Communication*. IEEE, 565–570.

[94] Rolf Pfeifer and Josh Bongard. 2006. *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press.

[95] Marc Pinski and Alexander Benlian. 2023. AI literacy-towards measuring human competency in artificial intelligence. In *Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS).*

[96] Jonathan A. Plucker, Ronald A. Beghetto, and Gayle T. Dow. 2004. Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist* 39, 2 (2004), 83–96.

[97] Posit team. 2023. *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA. Retrieved from http://www.posit.co/

[98] Jeb S. Puryear and Kristen N. Lamb. 2020. Defining creativity: How far have we come since Plucker, Beghetto, and Dow? *Creativity Research Journal* 32, 3 (2020), 206–214.

[99] R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[100] Mel Rhodes. 1961. An analysis of creativity. *The Phi Delta Kappan* 42, 7 (1961), 305–310.

[101] Mark A. Runco. 2004. Creativity. *Annual Review of Psychology* 55, 1 (2004), 657–687.

[102] Mark A. Runco and Garrett J. Jaeger. 2012. The standard definition of creativity. *Creativity Research Journal* 24, 1 (2012), 92–96.
[103] Jordan A. Sasser, Daniel S. McConnell, and Janan A. Smither. 2022. Investigation of aspects of embodiment on user perceptions in human-robot interactions. In *Proceedings of the 2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)*. IEEE, 1–2.
[104] Rob Saunders, Emma Chee, and Petra Gemeinboeck. 2013. Evaluating human-robot interaction with embodied creative systems. In *Proceedings of the 4th International Conference on Computational Creativity*. Citeseer, 205–209.
[105] Rob Saunders and Petra Gemeinboeck. 2014. Accomplice: Creative robotics and embodied computational creativity. In *Proceedings of the 50th Anniversary Convention of the AISB Symposium*, 1–4.
[106] Rob Saunders, Petra Gemeinboeck, Adrian Lombard, Dan Bourke, and Ahmet Baki Kocaballi. 2010. Curious Whispers: An embodied artificial creative system. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 100–109.
[107] Richard Savery, Lisa Zahray, and Gil Weinberg. 2018. Shimon the rapper: A real-time system for human-robot interactive rap battles. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 212–219.
[108] Stela H. Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E. Young. 2015. Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*, 125–132.
[109] Eva Specker. 2021. Further validating the VAIAK: Defining a psychometric model, configural measurement invariance, reliability, and practical guidelines. *Psychology of Aesthetics, Creativity, and the Arts* 18, 3 (2021), 449–462.
[110] Eva Specker, Michael Forster, Hanna Brinkmann, Jane Boddy, Matthew Pelowski, Raphael Rosenberg, and Helmut Leder. 2020. The Vienna Art Interest and Art Knowledge Questionnaire (VAIAK): A unified and validated measure of art interest and art knowledge. *Psychology of Aesthetics, Creativity, and the Arts* 14, 2 (2020), 172.
[111] James H. Steiger. 1980. Tests for comparing elements of a correlation matrix. *Psychological Bulletin* 87, 2 (1980), 245.
[112] Arthur Still and Mark d'Inverno. 2016. A history of creativity for future AI research. In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 147–154.
[113] Steven J. Stroessner and Jonathan Benitez. 2019. The social perception of humanoid and non-humanoid robots: Effects of gendered and machinelike features. *International Journal of Social Robotics* 11 (2019), 305–315.
[114] Patrick Tresset and Frederic Fol Leymarie. 2013. Portrait drawing by Paul the robot. *Computers & Graphics* 37, 5 (2013), 348–363.
[115] M. Alessandra Umilta', Cristina Berchio, Mariateresa Sestito, David Freedberg, and Vittorio Gallese. 2012. Abstract art and cortical motor activation: An EEG study. *Frontiers in Human Neuroscience* 6 (2012), 311.
[116] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS '17)*, 6000–6010.
[117] Veera Vimpari, Annakaisa Kultima, Perttu Hämäläinen, and Christian Guckelsberger. 2023. "An adapt-or-die type of situation": Perception, adoption, and use of text-to-image-generation AI by game industry professionals. *Proceedings of the ACM on Human-Computer Interaction* 7, CHI PLAY, Article 379 (Oct 2023), 34 pages. DOI: https://doi.org/10.1145/3549486
[118] Chetan Walia. 2019. A dynamic definition of creativity. *Creativity Research Journal* 31, 3 (2019), 237–247.
[119] Graham Wallas. 1926. *The Art of Thought*. Vol. 10. Harcourt, Brace.
[120] Bingcheng Wang, Pei-Luen Patrick Rau, and Tianyi Yuan. 2023. Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology* 42, 9 (2023), 1324–1337.
[121] Tom Ziemke. 2003. What's that thing called embodiment? In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 25. Psychology Press, 1305–1310.

Appendices

A Creativity Questionnaire

In both studies (Appendix B and Section 3), participants used visual analogue scales with gradations from 0 to 100 to evaluate the stimuli based on the following instructions:

Creativity. "Using your own subjective definition of creativity, the degree to which what you observe is creative".
Liking. "Your own subjective reaction to what you observe; the degree to which you like it".
Aesthetic Appeal. "In general, the degree to which what you observe is aesthetically appealing".

Originality. "The degree to which what you observe shows originality".
Technical Artistic Ability. "The degree to which what you observe demonstrates technical artistic ability".
Detail. "The amount of detail in what you observe".

Questions about creativity and its components such as Liking, Aesthetic Appeal and Detail were worded as close to Amabile's original questions [5] as possible, but we changed the object of evaluation from "the design" or "the work" to "what you observe". This change was made to avoid inducing a focus on the product, process or producer, but instead relate the questions flexibly to the type of PE which the participant is exposed to under different conditions. For Originality, we used similar wording to the previous dimensions to stay consistent. For Technical Goodness, we changed the wording from Amabile's "is good technically" [4] to "demonstrates technical artistic ability". We adopted this wording from Chen et al. [21] to ensure that the participants in the study do not confuse artistic skill with robot engineering.

Within CAT research, there is no consensus on the best scale to use and a variety of different scales has been explored, often without much justification [32, 79]. We chose to use a continuous, visual analogue scale for increased measurement granularity, and to ensure that the data can actually be considered continuous [61].

B Pre-Study

In order to assess differences in the perception of creativity in our study (Section 3), we had to assure that our drawings as stimuli had the potential to be perceived as creative in the first place. Moreover, to better isolate the effects of Embodiment and PE on the Creativity assessment in the study, we had to identify candidate stimuli that are, on average, perceived as similarly creative. We consequently conducted a within-subjects pre-study in which participants rated the Creativity of 18 candidate drawings, allowing us to eventually select seven drawings for the study. The study is in accordance with the Declaration of Helsinki, and was approved by Aalto University's Research Ethics Committee. The data used for analyses and other supplementary materials can be found in the OSF Supplementary.

B.1 Materials

As stimuli for both studies, we used human-made still life drawings of object arrangements. Examples of the drawings can be seen in Figure B1. Each arrangement consists of four objects, selected from three object categories: one artificial plant, one toy animal and two artificial fruits. The object categories are partly inspired by classic still life arrangements and comprise the following objects:

(1) Artificial Plants: A geranium, a monstera and an ipomoea.
(2) Toy Animals: A dinosaur, a bunny, a rat and a fish. The bunny is a plush toy, and the others are made of plastic.
(3) Artificial Fruits: An apple, a banana, a grape cluster, a kiwi, a lemon, an orange and a pear.

The arrangement of the objects is similar across all drawings; an artificial plant is in the centre and the rest of the objects are on both sides of it, a few centimetres to the front. The toy animal is always on the right side of the plant, facing a bit to the centre, and the two artificial fruits are on the left side of the plant, one of the fruits being a bit more on the centre and one a bit more to the left and back. The arrangement of objects was kept the same to prevent different arrangements from influencing the results, and to make them easier to recreate during the main experiment (Section 3). The overall drawing setup, including the position of the still life objects, light and the artist, was recorded to facilitate reproduction through the robots later on (see Figure B2).

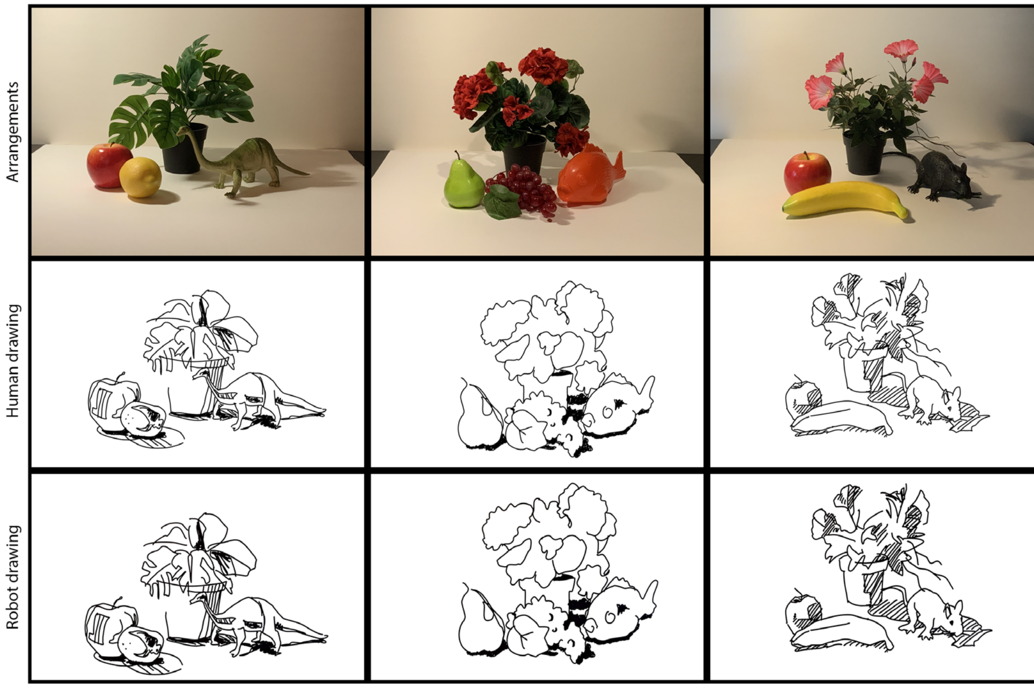


Fig. B1. Three examples of object arrangements (top), the corresponding drawings by the human artist (middle) and drawings recreated by the AxiDraw robot for the pre-study (bottom).

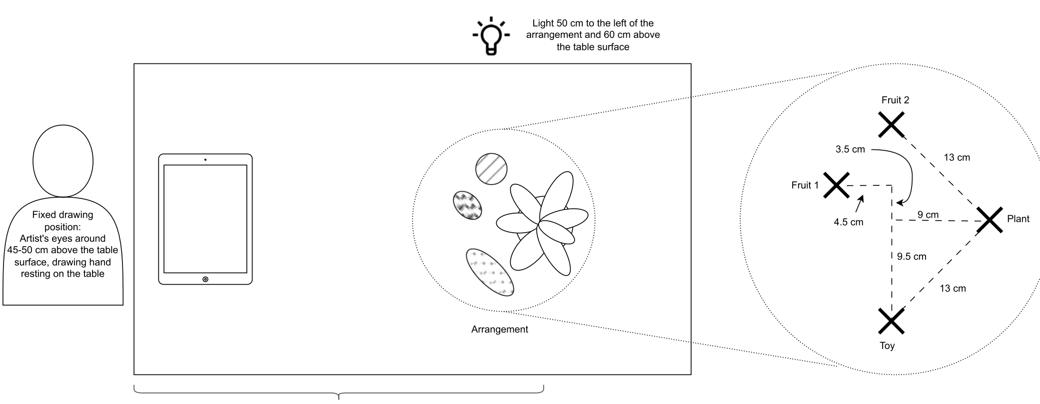


Fig. B2. Artist drawing setup and still life arrangement. The arrangement, consisting of four objects, is placed on a table, so that the centre object is one meter from the artist's table edge. It is lit from the side to create spot shadows. The artist rests their elbow on the table, thus keeping their posture nearly static during the drawing process. Their eyes were about 45–50 cm above the table surface.

We commissioned 36 drawings, each depicting a different combination of objects. The artist was instructed to use all combinations of plants and animals, but they were free to choose the fruits which best fit each arrangement, based on their artistic judgement. The commissioned artist has a Master's degree in art education and regularly attends live model drawing with strict time limits. Their active drawing time, excluding pauses of more than a few seconds, was limited to 5 minutes (±20 seconds). All drawings were drawn with a stylus in A4 landscape format on an iPad Pro 12+ using Affinity Designer 2. The drawings were exported as an SVG file with only SVG path primitives. The latter are read by the drawing software and converted sequentially into drawing commands for the robots to recreate the drawings. The robots recreate the lines in the same order as the artist originally drew them.

The drawings were reproduced by the robots on Fabriano Bristol smooth A4 marker paper[7] using a black Sakura Pigma Micron marker[8] with tip width 10. This pen and paper combination was selected because of its smooth finish as the friction when the pen moves on the paper is minimal, causing only minor artefacts to the drawings from the sometimes abrupt robot movements.

To avoid participant fatigue, we pruned this initial pool of recreated drawings to 19 instances which could be reproduced within a similar, limited time frame. This was done in three steps. Firstly, we replicated all man-made drawings on both robots and recorded the drawing times. The robots replicated the drawings at vastly different times compared to their human counterpart (who used 5 minutes ± 20 seconds of active drawing time for each), ranging from less than 1 minute to over 14 minutes. This is because the artist's hand movement speed varies differently during the drawing process compared to the robots' actuator. A detailed comparison revealed for instance that the artist took longer to draw intricate single lines, while the robots require more time for shading large patches. Secondly, we used these observations and asked the artist to redraw eight drawings, avoiding features that greatly slowed down the robot drawing performance, such as too much shading. Finally, we selected only those drawings which the slower of the two robots could replicate in approximately three to six minutes. This left us with the 19 drawings. Drawing times ranged from 215 to 324 seconds ($Mdn = 278$, $SD = 45$).

Inspecting the outcomes, we deemed the output of the different robots sufficiently close to not affect the assessment of their creativity (see Appendix C for comparisons); consequently, all 19 drawings used in the pre-study were drawn by the same robot (AxiDraw). The robots and their drawing setups are described as part of the study (Section 3). The final selection of the 19 drawings is provided in the OSF Supplementary.

B.2 Questionnaires

We used the same questionnaires for demographics, participants' expertise and interest information, and Creativity assessment as described in our main study (Section 3.3). The Godspeed questionnaire from the main study was not used, as participants did not encounter robots in the pre-study.

B.3 Participants

Thirty participants took part in the experiment during June 2023. Participants were recruited using mailing lists for Aalto University and University of Turku communities, and, to a much lesser extent, through the authors' professional and personal networks. Our study invitation (OSF Supplementary) encouraged people of varying levels of art expertise and background to participate in the study. Participants were required to have normal or corrected to normal vision, but this was not assessed in the study. Each participant received a 20€ gift card as compensation.

In response to a short demographics questionnaire, 13 participants identified as female and 17 as male. Their age ranged from 21 to 49 years ($M = 28.8$, $SD = 7.0$) and one participant chose not to specify their age. Thirteen participants reported having various kinds of art education (not limited to visual art), but these varied widely, with most reporting a singular or only a few courses of varying

[7] https://fabriano.com/en/product/bristol/
[8] https://www.sakuraofamerica.com/product/pigma-micron/

quality and length. Twelve reported having Master's, and seven to own Bachelor's degrees or equivalent as their highest completed degree, and the remaining participants reported high school or basic levels of education. Summaries of interest and expertise variables and comparison to main study participants can be found Appendix D. Each participant gave their written informed consent.

B.4 Procedure

The experiment was conducted at the premises of Aalto University and University of Turku, and took approximately 45–60 minutes to complete. Participants were offered to communicate with the experimenter in either English or Finnish, depending on their preference. They were first familiarised with the study through a written introduction, and then asked to provide written informed consent. As a deception, the introduction informs that the drawings presented in the study were drawn autonomously by a robot, taking approximately five minutes for each drawing. Participants were not given any additional information about the robot.

Next, they were shown an example drawing (Appendix B.1) along with a printed photo of the original arrangement represented by the drawing (for comparison, see Figure B1). This example drawing was separate from the 18 drawings to be evaluated, and the same for every participant. It was chosen as the median one in terms of robot drawing time from the 19 drawings. It served as a practice stimulus to explain the task, familiarise participants with the still life subject, and provide them with a baseline for the assessment of the remaining drawings.

For the core study, the participants were shown the 18 drawings and photos, one pair at a time, in two separate binders. The pairs were presented in a counterbalanced order to minimise order effects. After receiving each drawing, participants were asked to rate them with our CAT-inspired electronic questionnaire (Appendix B.2) on an iPad Pro 11. In accordance with the CAT technique, they were instructed to rate the drawings relative to each other.[9]

Once they concluded all ratings, participants were asked to provide demographics information and answer questions about their art interest as well as familiarity with AI and robotics, using the previously described questionnaires (Appendix B.2). Subsequently, participants were briefly interviewed about their overall experiences during the study, their personal definitions of creativity, and any other comments they wished to share regarding the experiment. Following the interviews, we revealed the deception that the drawings were originally created by a human artist and that the robot had only reproduced them from a script. We offered participants the option to withdraw their consent and data from the study based on this information. However, nobody chose this option.

B.5 Results

Overall, our drawings received a large variety of creativity ratings, with means ranging from 34.10 to 53.53 (overall $M = 46.93$, $SD = 20.83$). To estimate inter-rater reliability, we calculated a two-way random effects, average rating, consistency intraclass correlation coefficient (ICC), following the guidelines of Koo and Li [68]. Our ICC = 0.82 (95% CI = [0.67, 0.92]) indicates moderate to excellent reliability. While some drawings were perceived as less creative than others, these results suggest that none were perceived as uncreative, which would exclude them as eligible stimuli for our study. We present the creativity rating means and SDs of each individual drawing in Table B1.

To support our primary objective of selecting drawings that are similar in perceived Creativity, we eliminated variations in the participants' use of the rating scale by applying z-standardisation

[9]Amabile states in her CAT guidelines [5] that assessments should be conducted in a relativistic way, comparing items within a sample to one another. Only seeing one item at a time makes relative assessments harder, but it is necessary for us in the subsequent study (Section 3) in order to not leak between accumulative types of PE, so we adopted the same process here to obtain a consistent rating procedure between studies.

Table B1. Mean Perceived Creativity Scores and SDs for Each Drawing

Drawing	Mean	SD	Drawing	Mean	SD	Drawing	Mean	SD
1	53.53	23.89	7	49.10	21.91	13	44.00	20.86
2	51.80	21.93	8	56.43	21.64	14	46.73	20.11
3	55.23	21.87	9	51.37	21.07	15	42.60	18.80
4	50.20	20.43	10	44.67	20.70	16	49.63	20.85
5	43.27	19.35	11	50.50	20.78	17	35.27	15.07
6	43.53	19.81	12	42.80	17.72	18	34.10	15.86

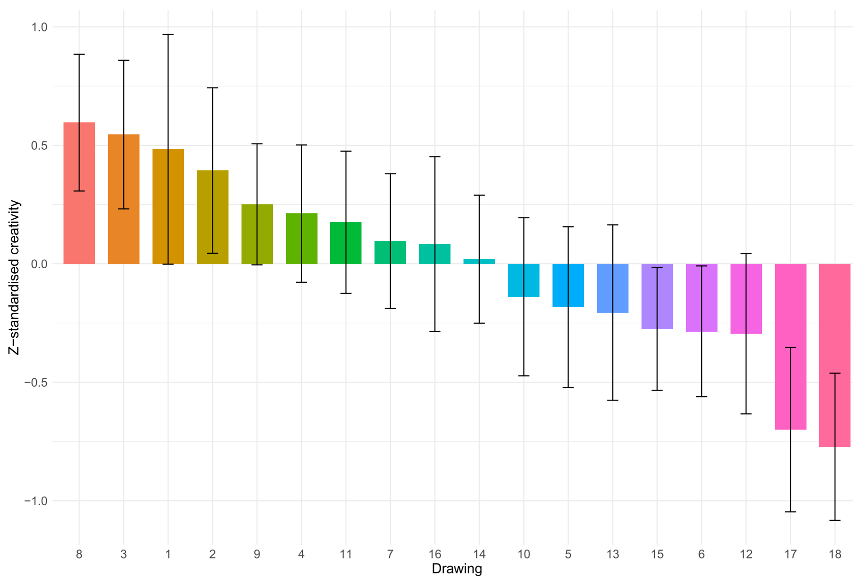


Fig. B3. Means of z-standardised perceived creativity scores with 95% confidence interval for each drawing.

to the ratings. The mean z-scores of perceived creativity and 95% confidence interval are presented in Figure B3.

Drawing Selection. In selecting the drawings, we aimed to exclude outliers and drawings at the extreme ends of the rating scale. We also avoided selecting drawings that were very similar in terms of combined object arrangement and artistic style, to ensure diversity in the stimuli for our study. Consequently, we chose the following drawings, which are consecutive in ratings of Creativity and lean towards the higher end of the rating interval without reaching the extremes: 4, 7, 9, 10, 11, 14 and 16. This final selection is shown in Figure 1.

Exploratory Analysis. We explored the relationships between the perceived Creativity and other items on the creativity questionnaire (Appendix B.2: Liking, Aesthetic Appeal, Originality, Technical Artistic Ability, Detail) to understand whether participants could differentiate between them. Moreover, given that these have been frequently reported as factors in the creativity assessment [102] but are sensitive to the assessment context, we wanted to identify their role in the overall creativity assessment in this particular scenario. To both ends, we calculated correlations between all rated properties for each participant and averaged them into the matrix in Figure B4.

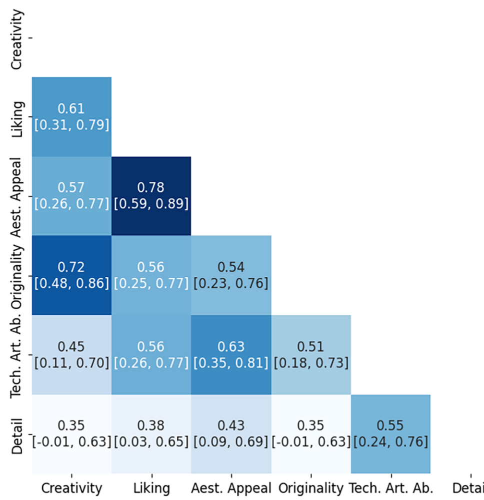


Fig. B4. Averaged correlation matrix in a heatmap form for assessed drawing properties from the pre-study. Cells include Pearson correlation coefficients and their 95% confidence intervals. Darker shades indicate higher correlations.

While there are strong connections between assessed properties, overall participants seemed to be able to mostly differentiate between them. We found that Originality was most strongly associated with the Creativity ratings. In contrast, Liking and Aesthetic Appeal were moderately related to Creativity, and Technical Artistic Ability as well as Detail were only weakly related to Creativity. The highest sub-item correlation was observed between Liking and Aesthetic Appeal, whereas Technical Artistic Ability and Detail were only moderately or weakly correlated to the other responses. Overall, all of the responses were at least somewhat correlated with each other.

Interview Feedback. Our interviews supported these findings; in particular, participants emphasised the challenge of distinguishing between Originality and Creativity. Most defined creativity as the act of adding, changing, highlighting, or omitting some elements from the original arrangement. Originality was often judged in a similar vein, although encountering something novel, like a unique drawing style, was reported as having an influence on its assessment.

Further supporting our quantitative findings, the vast majority of participants deemed the drawings overall creative, viewing them as distinct interpretations of the arrangements rather than mere replications. However, some expressed the belief that portrayal art is inherently not very creative, as they miss a stronger role of the artist's imagination.

Finally, participants highlighted the assessment of the initial drawings relative to others as challenging, in that they missed a stronger comparison baseline. Based on this feedback, we added an additional example drawing in our subsequent study, as detailed in Section 3.2.

C Robots

AxiDraw is a robotic pen plotter, commercially produced for drawing and writing purposes. It is equipped with two precision stepper motors in its base that drive a single looped belt, controlling the carriage's movement around in the XY plane. A third smaller motor is used for lifting up the pen off the page. While operating, AxiDraw produces distinctive mechanical high-pitched sounds. The marker was attached as the robot's end effector using its original pen holder in upright position (tip is in 90 degree angle to the paper). Rubber band was used to give spring to the pen holder to

Fig. C1. Example of the differences in overlaid drawings produced by AxiDraw (pink) and xArm (blue).

keep the pen on the paper surface. With AxiDraw, we use the API provided by the manufacturer[10] to convert the SVG file of the original drawing into drawing commands.

In contrast, xArm is a general-purpose shiny white robotic arm with metallic colouring near the joints. It is powered by an industry-grade harmonic drive and servomotors. With six base and rotary joints, it provides six degrees of freedom. For xArm to hold the marker, we 3D-printed a modified model of a pen holder end effector, consisting of two parts: an outer part that is attached to the xArm's flange, and an inner part which holds the pen and is attached to the main part with rubber bands to obtain the right spring for the pen. To convert the SVG file of the original drawing into drawing commands, we use the Pilz Industrial Motion Planner and KDL Kinematics Plugin, both provided by the MoveIt-package for ROS2 Humble.[11]

We denoted a mean ratio of 2.30 for the drawing times between the robots, AxiDraw being faster than xArm. As xArm was already using its maximum acceleration and speed, we adjusted AxiDraw's drawing process to be approximately as slow as xArm's to accommodate for similar time periods of observation during the main experiment.

In order to assess the similarity between drawings produced by the two robots, we repeated the same drawing five times for each robot, yielding a total of 10 drawings. To quantitatively compare the drawings, we employed two metrics: the Structural Similarity Index (SSIM) and the Mean Squared Error (MSE). SSIM measures the structural similarity of two images by taking into account the luminance, contrast and structure of the images. It has a value range of -1 to 1: a score of 1 indicates perfect similarity, 0 indicates no similarity and -1 indicates perfect anti-correlation. On the other hand, MSE measures the mean squared difference between the images. MSE values range from 0 to infinity: a score of 0 indicates identical images, while higher values indicate greater dissimilarity.

For the drawings produced by the two robots, the MSE was 2,894.36 and the SSIM was 0.91. A visual comparison of two drawings is presented in Figure C1. We argue that the minor variations in drawings are not enough to betray which robot made them, particularly when participants are only exposed to them one at a time.

[10]https://axidraw.com/doc/py_api/
[11]https://moveit.ros.org/

D Interest and Expertise Comparisons

The interest and expertise variables from both studies are reported in Table D1. To investigate whether our participants differed in terms of these characteristics, pairwise comparisons, also reported in the Table D1, were conducted. Assumptions for parametric *t*-test were not met, so Mann–Whitney *U* tests were used. After adjusting for multiple comparisons using Holm–Bonferroni method, we did not find any statistically significant differences between the two groups for any variable. This suggests that our pre-study and main study participants were similar in terms of Art Interest, Art Practise, AI Technology Knowledge, AI Usage Experience and Robotics Experience.

Table D1. Means, SDs and Pairwise Comparisons using Mann–Whitney *U* Test of the Interest and Expertise Variables in the Pre-Study and Main Study

	Pre-Study (SD)	Main Study (SD)	U-Value	Adjusted p-Value
Art Interest	48.47 (11.81)	49.3 (12.4)	926	1
Art Practice	3.70 (1.94)	3.35 (2.18)	797	1
AI Technology Knowledge	13.90 (5.06)	14.4 (4.42)	937	1
AI Usage Experience	9.50 (3.45)	10.4 (2.91)	1,030	1
Robotics Experience	5.67 (2.96)	7.07 (3.06)	1,141	0.19

p-Values are adjusted using the Holm–Bonferroni method.

Received 23 May 2024; revised 16 October 2024; accepted 11 December 2024