The Eyes, the Hands and the Brain: What can Text-to-Image Models Offer for Game Design and Visual Creativity?

Hongwei Zhou
University of California, Santa Cruz
Santa Cruz, USA
hzhou55@ucsc.edu

Jichen Zhu
IT University of Copenhagen
Copenhagen, Denmark
jicz@itu.dk

Michael Mateas
University of California, Santa Cruz
Santa Cruz, USA
michaelm@soe.ucsc.edu

Noah Wardrip-Fruin
University of California, Santa Cruz
Santa Cruz, USA
nwardrip@ucsc.edu

ABSTRACT

Text-to-image models such as DALL-E, Stable Diffusion, and Midjourney have seen a boom in development and adoption in both commercial and hobbyist spaces. This paper is a theoretical analysis aimed at informing the development of games that help improve critical literacy around text-to-image models. It asks: what assumptions and perspectives do text-to-image models have on visual creativity, and how do we bring that out through games? We propose a theory to differentiate between seeing an image through the expression of color, shapes and lines, and seeing an image through the recognition of concepts and ideas. These two ways of seeing are two different ways of orienting the player/user to their visual creativity. While traditional painting mechanics emphasize the former, text-to-image interfaces emphasize the latter. We deploy this perspective to study games with traditional painting interactions and games with text-to-image interactions. This paper hopes to contribute to design both broadly for games about visual creativity, and narrowly for gameplay with text-to-image models — specifically, how the latter fosters a different type of visual creativity than traditional painting interactions.

CCS CONCEPTS

• **Human-centered computing → Interaction design theory, concepts and paradigms**.

KEYWORDS

Game Design, Text-to-Image, Painting, Visual Creativity, Stable Diffusion

ACM Reference Format:
Hongwei Zhou, Jichen Zhu, Michael Mateas, and Noah Wardrip-Fruin. 2024. The Eyes, the Hands and the Brain: What can Text-to-Image Models Offer for Game Design and Visual Creativity?. In *Proceedings of the 19th International Conference on the Foundations of Digital Games (FDG 2024), May 21–24, 2024, Worcester, MA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3649921.3650001

FDG 2024, May 21–24, 2024, Worcester, MA, USA

ACM ISBN 979-8-4007-0955-5/24/05.
https://doi.org/10.1145/3649921.3650001

1 INTRODUCTION

There is little doubt that creative tools play a major role in how one engages with creative acts. The recent boom of text-to-image machine learning models has struck the visual creative communities because the models are offering a radically new modality to create aesthetic images. The capability of utilizing natural language interfaces raises much discussion and debate on the nature of art, creativity, as well as the rights and working conditions of visual artists. Besides these pertinent ethical issues, what this new interactive modality offers for user/player experience remains an open question. Similarly, how this new modality can open up new possibilities for video games also remains largely unexplored.

We situate our work within the economic and cultural instability that text-to-image models bring to the visual creativity landscape. The goal is to communicate these questions through the medium of games. Specifically, this paper is a theoretical starting point to develop games that seek to foster critical literacy around text-to-image models. Critical literacy means that one shouldn't take the text-to-image model for granted, in that it shouldn't fall into a naive optimism of AI for its efficiency and "democratization" of visual creativity. Rather, text-to-image models enable a particular attitude towards visual creativity, promote certain ways of knowing and certain ways of creating that are very different from traditional visual art making like painting.

This work attempts to develop a theoretical grounding for such criticality and also to inform game design decisions so that such criticality becomes explicit. The critical literacy we seek to develop here is different from the traditional motivation of theories around literacy — education [25]. Here, we position our intervention within the larger discursive ecology around AI and text-to-image models. Thus, we target a general audience with some awareness of the current AI discursive landscape. In the words of literacy theorist James Paul Gee, we want to make sure that the meta-level cognitive and linguistic skills developed via the game can be used to critique the discourses around AI [24]. Therefore, we do not see the games we hope to make as merely feeding information to the players, but rather presenting our particular arguments and discoveries about certain ways of interpreting the issues around text-to-image models, in the hope of generating more discussion and thinking.

We attempt to address two questions in order to inform a critical game design approach. First, how do we understand text-to-image model's relationship to visual creativity? Visual creativity here means the creative act that a person engages in to produce images to communicate ideas and intentions. To answer this question, we are inspired by art historian Aloïs Riegl and French philosopher Gilles Deleuze's discussion of art theory and history. Riegl and Deleuze develop a metaphorical way to theorize how the creative apparatus engages with the activity of painting. This creative apparatus consists of different organs of the artist's body such as the eyes and the hands, and how a particular approach/technique to painting signifies a particular relation of these organs (the hands must replicate the curve that the eyes see in realism vs. freely expressive hands and bodies in abstract expressionism). Although this description of organ is more metaphorical rather than physiologically accurate, the main theoretical consideration is how the player/user subject becomes oriented towards a certain way of expression/creativity. In Section 3, we propose that text-to-image models promote a particular way that the brain and the eyes are related to each other; instead of seeing visual elements like shapes, color and compositions as expressive of ideas in themselves, seeing becomes more like recognizing objects and concepts where shapes and colors are subsumed to the function of depiction.

Second, what are some game design features used to structure how players express their visual creativity? Because games about creativity can offer too large of an action space to infer possible play experience, we pay close attention to how each design feature constrains and limits creative approaches, or simply just makes particular approaches more obvious and easier. In Section 4, we study the design of two painting games: *Passpartout 2: The Lost Artist* [23] and *Chicory: A Colorful Tale* [35]. We define painting games as video games that center on the pleasure of painting while still providing a gamified experience. This is a nuanced definition as they are distinct from painting tools (not enough game) and games with painting mechanics (do not center on the pleasure of painting). We identify design features in these two games that orient the player to be visually creative. We also study experimental games that utilize text-to-image models in their mechanics: *Bureau of Multiverse Arbitration* (BoMA) [1], *Keyo Against Humanity* [9] and *Laughprop* [57], to also show design decisions involved to promote creativity. This small survey serves to show how games, unlike painting tools, can utilize different design features to assist and bootstrap players' creative engagements. Since painting interfaces and text-to-image interfaces are radically different ways to produce aesthetic images, identifying different design features along this distinction helps show different solutions to structure the player's creative agency.

The player/user has to express themselves based on what is provided to them; their expressions are also structured by the tools in terms of affordance, constraints and limitations. This is where we see the two questions connect: in the same way that text-to-image models would structure visual creativity (how one relates to their creations) differently, the game design features on creativity would also structure the ways that the players express themselves. In Section 5, we apply our theories and surveys to inform our ideation for potential design elements in a critical game. While mostly utilizing our theories and light surveys for game design purposes, we hope to have offered theories and design lessons to bootstrap further studies of visual creativity in games, and the possibilities of text-to-image models for games.

2 RELATED WORK

2.1 Text-to-Image Models and Creativity

This subsection explores discussions around AI and visual creativity, especially recent machine learning-based image generation, and how they raise new questions around art, aesthetics, and philosophy. To start, the concept of creativity itself has a rather divergent set of definitions. It shifts and mutates based on which domain (social, cultural, cognitive etc.) the scholar wants to work with. We here offer two questions to specify what we mean when creativity is mentioned in this paper: how does creativity emerge? and how is creativity recognized? The first question touches on how scholars examine different phenomena to study the emergence of creativity. Notably, much creativity scholarship defines different phases of creative acts such as ideation, creation, and verification to understand the back-and-forth cognitive processes [22, 59, 63]. Other theories, such as Rhodes' four P model and Amabile's management study think about the larger environments that contribute to a person's creative acts [3, 6]. Here we're interested in a particular factor in organizing creative acts: the tools, in order to study how text-to-image models can shape visual creativity in different ways. The second question on the recognition of creativity also differs by scholarly context. A high-level divide can be drawn between the scale of the personal and the scale of the societal, cultural, and historical, such as Boden's "H-creativity" (historical) and "P-creativity" (psychological) [6, 14]. Different scholars expand into each scale in different ways. For example, Amabile defines creative ideas as novel and useful, mostly grounding them in a kind of social recognition, while Kaufman & Beghetto's four-C model expands upon the P-creativity, or the Little C, to talk about ways an individual can be creative in their day-to-day lives [33]. We're particularly interested in creativity as a personal act because we are interested in the experience of use and play. As Compton describes, this perspective focuses more on the psychological connection to one's creation, where creativity is less about looking at the product rather than production itself [12]. Specifically, we are interested in looking at how a player/user comes to see object imprints, be it paint on paper or traces on a digital canvas, as *expressive of* some ideas or intentions of the player/user. Visual creativity, then, means the creative act that a person engages in to produce images to communicate ideas and intentions.

Defining creative acts as being expressive of ideas on a personal level allows us to connect to a body of literature around artificial intelligence and visual creativity, as the literature centers around the question of *authoring through AI*. Boden & Edmonds have written extensively about different types of computer art and explore questions around aesthetics [7]. They highlight the question of artistic authorship and agency as the key question around changes in aesthetic criteria; which part of the tool allows the authors to express ideas also determines how one would appreciate the artwork itself. The thread of authorship underlies Audry's book on art and AI, which walks through different processes of machine learning techniques to point out ways artists can ask questions and

offer artistic solutions [5]. Sivertsen et al. analyzed a collection of AI art and traced creativity to how artists utilize the underlying ML process to create meaning [55]. Celtnic & She review discussions around generative image AI and observe that neither disruptive technology nor the question of authorship is new to art, drawing connection to photography and action paintings of Jackson Pollock [10]. What should be emphasized, then, is avoiding thinking of generative AI as somehow autonomous and instead keeping in mind the human involvement in the creation of aesthetic images. Oppenlaender borrows from Rhode's four P creativity to argue that one shouldn't assert the presence of creativity simply by looking at the product. Rather, one should look at the *process* and *press* to understand how human creativity happens with text-to-image models. The process includes iterative prompt engineering and image curation as ways humans exert their intentions. The *press* includes communities that share prompting knowledge and appreciate the images [45]. We differ from this scholarship in that we focus on the very modality of the interaction itself: text-to-image, and theorize about how this modality of interaction shifts the way the subject relates to their outputs.

Philosophically, we align with Zylinska's post-humanist approach to authorship in the light of AI art [66]. Borrowing from various philosophers, Zylinska argues that creative acts always occur within a larger apparatus that includes cultural and technological factors. Technology here has a more general sense as objects of *techniques*, which even includes human organs such as the eyes and the hands. This post-humanist view frames agency as distributed in human and nonhuman actors. The humanist notion of an identifiable and centralized creative author is thus a rather misguided notion. Instead, Zylinska wants to refocus on the cultural, social, and economic conditions that maintain themselves around particular humanist notions of agency, which AI is set to challenge. It is from this perspective that we motivate our work: that creativity takes place in a *distribution of authorship and agency*, where each part of the apparatus can condition other parts. We seek to theorize about how text-to-image models might condition the interacting subject in certain ways, and how this conditioning might habituate certain thoughts and actions. We intend to motivate game design decisions to make this conditioning from the tool explicit. Video games are appropriate to demonstrate this issue because, as philosopher Thi Nguyen argues, they work in the medium of agency [44].

2.2 Creativity and Design

We also relate this literature to two different scholarly fields in design. First, the game design literature offers formalist approaches to considering the study of game structures and user experiences [13, 50, 65]. They are formalist because they mostly study the forms of games in terms of their rules, incentive structures, and interactive dynamics to infer player experience. While this paper hopes to contribute to the game design literature, it diverges in a significant way that is yet to be crystallized in the field: a focus on games that provide significant creative freedom to the players. One difference we locate is that formalist game design approaches are typically studied through what kind of narrative and interactive affordance is provided by the game, or in terms of positive action spaces. However, when studying player creativity, one typically sees a very large action space where player behaviors vary too much to be described effectively. A surface-level understanding often connects creativity to a game's open and sandbox nature, in other words, a large action space [26]. However, to our knowledge, not much has been written about how game design is leveraged to foster creativity within the player, especially not in a way that is specific to a type of creativity. Literature that connects creativity to video games is often from a cognitive science context with studies that do not intend to contribute to design [30, 47]. (Rahimi & Shute provide some design suggestions from an educational perspective, such as providing examples and guides to kickstart a student's creativity [48].)

One popular design theory that can be connected to creativity is flow. Flow denotes a state of hyper-focus where the player is completely immersed within the game activity [14]. Design approaches such as clear goals, immediate feedback, and balance between challenge and skills are often cited for inducing flow state [11, 47, 53]. While we see our contribution to design as creating opportunities for flow, we do not inherit the design approaches mentioned because they are somewhat high-level and non-specific to game design elements, and also too specific in terms of design philosophy to immerse the player subject, which we see as not quite suited for developing critical literacy. Overall, we are interested in how specific game design elements can orient the player in their creative process. We conduct our game design survey in a way that pays special attention to how each game design decision *limits* and *constrains* the large action/possibility space, or simply makes certain actions more obvious or intuitive.

The second design adjacent field is the study of Creativity Support Tools (CSTs) within the field of Human-Computer Interaction (HCI). The field is mostly intended to provide digital aids, or in the words of Shneiderman, "to make more people more creative more often" [52]. Similar to how creativity research lays out different phases within a creative act, Shneiderman also lays out many basic activities that relate to fostering creativity: to learn from examples, to receive feedback through socializing, to create, and to disseminate [51]. Much of the HCI literature is inspired by creativity research from the perspective of cognitive science [22]. Creativity support tools are thus often positioned as integrated into these cognitive processes (often involving problem finding, ideation, creation, and evaluation [49]), lowering their barriers. We differ also from this field of research in that we do not attempt to identify a particular problem within creative acts in order to suggest a technological solution. Instead, we consider the technology *as problems in of themselves*, problems in this sense mean sources for thinking, contemplation, and expression. This is similar to Li et al.'s alternative approach to CSTs by introducing the notion of power from a more socially critical perspective. Instead of focusing solely on user empowerment, the tools themselves also exhibit the power to "mediate how users work, and designing a tool means structuring and bounding its users' ideas, goals, and intentions" [34].

3 A HUMANISTIC FRAMEWORK ON TEXT-TO-IMAGE

This section presents our extension of theories of art to analyze text-to-image art production. This theoretical work will not encompass the entire influence that current AI techniques will have on artistic production, as new ones are invented quickly and are still being adopted practically. The theory will simply account for the very specific modality of text-to-image interaction, as in, using natural language interfaces to generate aesthetic images. Based on philosopher Gilles Deleuze's theories on painting as inspired by Aloïs Riegl [15], this section will examine how the creative process of art changes with text-to-image interaction.

At first glance, it might be difficult to see how a theory developed for high art can be applied to game analysis and human-computer interaction. The key connective tissue is the *structuring of creative agency*. The first subsection will discuss art movements, styles, and techniques that structure and limit possible expressions in painting. It is through contrasting different art movements and styles that this structuration is revealed. In the same way, as discussed in Section 3.2, the tools will structure the artist's agency in various ways. Although very different sources of influence, they all come together to constrain artistic expressions while also making them possible in the first place.

3.1 Painting as Organizing the Organs

We theorize about the creative process of painting by building an image of the relationship of organs: how different parts of the artist's body relate to each other during painting. The starting point of this idea is to not assume a universal and static being of the human body; the human body is dynamic, constantly in flux, and contingent on a continuous being with the world.

Another way to understand this is how perceiving art is inherently *selective*: a viewer would not attribute an equal amount of attention to every part of the painting. Rather, different parts of the painting are highlighted based on one's knowledge, experience, and even mood. Even the distance of viewing changes how the viewer would perceive a painting. This selective nature of viewing art also applies to creative processes, where the artist would perceive their own work in different ways in different stages of creation. The eyes and the hands would relate differently when outlining, tracing, shading, or texturing. Sometimes even just clearing one's head and sitting back to re-examine the work would create a different understanding of one's own work.

We discuss different ways the brain, the eyes, and the hands can relate to each other in painting, and then apply this perspective to examine interacting with text-to-image models. It is not a scientifically accurate description of how the human body changes, but rather a metaphorical way to talk about painting that brings one closer to the dynamics of the creative process.

The art historian Aloïs Riegl originally developed this way of examining different genres of art and painting through the relations between human organs. One example is his distinction between "close-range vision" and "far-range vision" [60]. The former describes a way of seeing where one's eyes are imagined to be so close to the subject that seeing inevitably invokes the sensation of touch. Seeing is more about comprehending a subject's contours rather than its colors and shadows. This way of seeing is more relevant to art forms such as sculpture. The latter is more related to classical paintings, where the eyes have to step back to analyze light and shadows, and the separation between individual objects and parts is made through the perception of depth. This simple example shows how ways of seeing can vary depending on what the artists are working on.

The late 20th-century philosopher Gilles Deleuze extends the way Riegl talks about the painting process. To Deleuze, a classical or renaissance painting is only a particular way to relate the eyes to the hands, where the eyes analyze the colors and shadows while the hands and the brushes replicate those observations[1] [15, p101]. But modern art shows different ways that eyes and hands can relate. One example is impressionist or post-impressionist paintings such as the one on the left of Figure 1, whose brush marks are very visible because the artist sacrificed accuracy in forms (shapes and contours) in order to capture the lighting conditions quickly. Abstract expressionism, as exemplified by Jack Pollock in the middle of Figure 1, shows what Deleuze calls "manual" [15, p124], where the hands and brushes are "freed" from the eyes to do away with consistent shapes and real objects. The eyes are merely evaluators of the result of the improvisational hands. On the other hand, the eyes can also completely dominate the painting to a point where the hands are almost a nuisance, such as Piet Mondrian's painting on the right of Figure 1, which is about compositing with pure visual forms. Deleuze would call the hand "digital" in this style called minimalism, as the image seems to be better represented in digital form, where the "hands" that paint would become almost mechanical.

There is a parallel between Deleuze's "digital" and how the hands are used in text-to-image models, where the hands are "reduced to the finger" by typing away on the keyboard [15, p124]. The marks of the hand can no longer be directly read on the image like the Monet's painting in Figure 1. Rather, examining a text-to-image result up close only reveals the limit of the image model, where the edges of different objects get confused and the lines are no longer consistent [42].

Although these discussions of how the hands and eyes are related are mostly about studying different artistic techniques, styles, and ideas, we find this framework beneficial for thinking about how the player/user subject might orient themselves to a creative tool/game. The contrast between using the hand to replicate information from the eye vs. using the hand in improvisational painting illuminates different ways the eyes and the hands are organized and thus afford/constrain certain types of expression and meaning-making.

Different implementations of painting in a tool or a game also orient the subject to their creation in different ways. For example, the way a digital brush is set up without pressure simulation might not emphasize the movement of the hand in certain ways. Or the fact that digital painting has to be rendered through the screen diminishes certain visual ideas in the materiality of the work. In our study of painting game design features in section 4, we frame certain features by how they constrain certain ways the subject's body can express itself on the canvas.

[1] This is a simplified characterization that is mostly set up as a contrast to other painting styles in modern art, as discussed later.

Figure 1: Examples of modern paintings. From left to right: *Rouen Cathedral, West Façade* by Claude Monet [41], *Number 1A, 1948* by Jackson Pollock [46], *Composition with Red, Blue and Yellow* by Piet Mondrian [40]

3.2 Two Ways of Seeing

In addition to the eyes and the hands, we want to introduce another human organ: the brain. Here, the brain means understanding, intellectualizing, and ideating as a part of the viewing and creating process. What am I drawing? What is happening in the painting? What are the objects in the painting? The brain has a complicated role in both viewing and creating paintings. Gilles Deleuze goes as far as saying that the goal of painting is to go beyond comprehension in the brain, or the "figurative, narrative and illustrative" [15, p6]; the main takeaway of a painting cannot simply just be recognition of objects and narrative events, but rather something that can only be communicated through painting as a visual medium. This criterion is perhaps too harsh, but it is valuable to consider what information painting is communicating outside of the depiction of objects: the colors, the subject's poses, and the perspectives are all relevant factors for communication.

For Deleuze, ideas in the brain are considered drawing in their most primitive and undeveloped form. He argues that it is a mistake to say that a painter has the most agency when presented with an empty canvas. Rather, their brains are already filled with *clichés* - "ready-made perceptions, memories, phantasms" [15, p71]. The artists have to work against expectations and habits to discover a unique work. In this way, the painting process is almost like subtracting and discovering, where the painter thinks *with* the painting as the artwork progresses.

To further this characterization of the brain, much visual art training also involves unlearning by actively working against habits in the brain. One of the first exercises in *Drawing on the Right Side of the Brain* is to replicate a line drawing by looking at it upside down, so the drawer cannot discern what each line is supposed to represent in relation to the object [19] (as demonstrated in Figure 2). This is also mentioned by Zeller, who talks about "blocking in" as a process where an artist thinks in shapes to "sidestep the distortions that happen when your mind works symbolically" [62, p100]. Cartoonist Scott McCloud also draws the distinction between resemblance and the picture plane. The former considers an image's resemblance to real-world objects, while the latter works on the level of abstract shapes [38]. In these accounts, recognition of objects in the brain almost dominates how the eyes see, in a way that the eyes no longer "see" what they actually see. By looking at an image upside down, the eyes are finally free to see lines in their actual relationships.

Overall, we want to draw a distinction — between *seeing to perceive visual information and communication* and *seeing to recognize ideas and objects* — and use this distinction to analyze text-to-image models. The latter is an untrained brain conditioning the ways that the eyes can see, while the former is the brain acknowledging visual information as its own way of thinking and communicating. Another way to frame this distinction is that one works with an ontology/level of abstraction wholly different from the other. In the former, the artist works with color, shape, lines, compositions, and materiality directly, and thus, they can define their expressions within the manipulation of those visual units. In the latter, a text-to-image prompter works with descriptions, ideas, and objects mediated within natural language, where the expressiveness of color, shapes, and lines are subsumed/limited into automation and the realization of the text prompt. In a way, colors, shapes, and lines are "freer" to become expressive in the former than in the latter. While theoretically a perfect text-to-image model can certainly follow descriptions so exactly that these two ontologies collide, it is not what text-to-image models are made to easily do.

Following this distinction, we can say that text-to-image models promote and habituate the latter way of seeing: seeing as recognition. There are many factors that reinforce this way of seeing as the primary mode of perceiving text-to-image outputs.

(1) The author's interaction with images is driven through natural language. The intention of the author has to be explicitly and externally inscribed through text. In a way, seeing as recognition is even more encouraged because text is so ever-present in the interface and such a core part of the interaction loop.
(2) The large volume of images generated also pushes one to default to the recognition of ideas in order to grasp the variations. This can be seen in many instances when AI-generated images catch popular attention. Trends like Wes Anderson-style trailers and Harry Potter Balenciaga memes quickly led

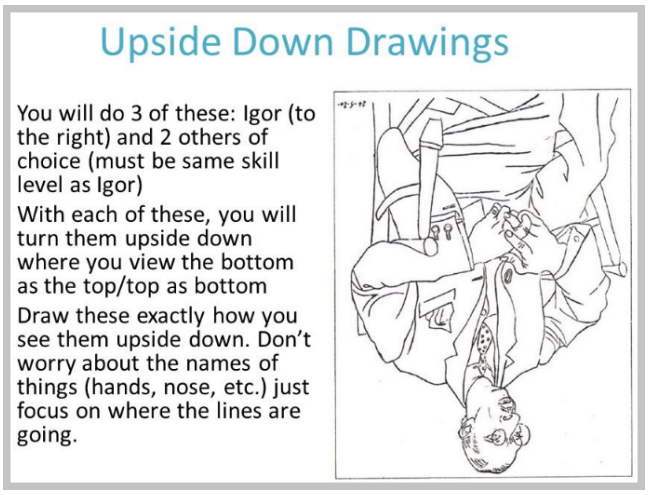

Figure 2: Upside down drawing exercise from *Drawing on the Right Side of the Brain* [19], slideshow taken from [17]

.

to a great number of variations on similar ideas [31]. But the variations take place on the level of cycling through recognizable brands and ideas (Lord of the Ring in Wes Anderson style, Breaking Bad Balenciaga etc.).

(3) This focus on recognition is further emphasized by current text-to-image models' difficulty in tasks such as human anatomy and object position descriptions — in other words, how successfully the models can realize textual descriptions. Much technical research and prompting effort focuses on eliminating such "mistakes" that hinder the recognition of objects [32, 37].

An artistic sensibility on visual communication becomes an optional trait in text-to-image production and appreciation, as artists can often take on the role of a curator, and an appreciation of the generated result does not go beyond a vague impression of "quality." (It is thus not surprising that Sivertsen et al. found that contemporary AI artists often intentionally evoke ambiguity [55] in order to challenge the fixed ways of seeing as visual communication as well as seeing as recognition.) As AI art is still in its infancy, the appreciation of such quality is often directed towards appreciation of the capability of AI in general, fueling the AI hype as part of its artistic message.

A recognition of ideas and objects as conditioned by text is also a very specific type of recognition, as text has its own way of drawing boundaries, marking sameness and differences. The available vocabularies for emotions in the dataset limit the possible ways that generated images can communicate affect. The notion of "style" is a clear illustration of this limit of natural language. As text-to-image prompts often separate the subjects and the visual styles of the image ("War, in the style of Claude Monet"). But a cursory understanding of art history would show that these two are not clearly separable in defining the style of an art movement [27]. The recent mass production of AI images in the style of Wes Anderson often misses how Wes Anderson's style is defined by how his storytelling, themes, characters, and film-making all complement the visual style to form a holistic character of Wes Anderson as an artist [18]. Here, the medium has become the message, where the meaning of style is already biased by how the technology is constructed.

4 ANALYZING VISUAL CREATIVITY IN GAMES

The main contribution of the previous section is the distinction between *seeing as visual communication* and *seeing as recognition*. We argue that the distinction between the two is realized in the two different ways to approach visual creativity: working with lines, colors and shapes directly vs. working with text-to-image. These two different interfaces work with different ontologies/levels of abstraction to produce aesthetic images. This section examines games with these two different interfaces to study how game design decisions come to structure the creative processes built on these two different ontologies. We want to study these design decisions with special attention to how they structure the player's creative agency by limiting, constraining and highlighting certain approaches when encountering a large space of possibilities.

4.1 Design Features of Commercial Painting Games

This subsection closely studies the design of two painting games, *Passpartout 2: The Lost Artist* and *Chicory: A Colorful Tale*, to demonstrate game design strategies for fostering visual creativity. A painting game is a delicate balance between

- *Bringing visual creativity, specifically painting, to the foreground of player interaction.* The player should actively consider visual language aspects such as colors, lines, compositions, forms, and textures etc. *for their own sake* rather than instrumentally for other goals. The game should either explicitly (through rules) or implicitly (through expectations set up by framing, genre, or advertisement) encourage these considerations from the player, and
- *Being somewhat game-y.* This can mean various things, and we do not want to impose a fixed definition of what it means to be a game. But some features that qualify as game-y would include fictionalization, additional incentive structures, or a rule set outside the pure act of painting.

We do not consider games with an obvious imbalance towards one another to qualify as painting games. For example, games such as *SuchArt* [39] and *Vermillion* [58] are mostly about simulating the painting experience in a virtual space, which falls short on the game-y criteria.

On the other hand, many games with some semblance of painting mechanics are better described as games *with* painting rather than games *about* painting. In other words, they have become too game-y. For example, the game *INK* [61] is a 2D platformer where the player and enemies can spray various paint colors on the platforms, creating a unique visual flair enabled by the platforming mechanics. However, *INK* clearly doesn't fall under our consideration of painting games mainly because the incentive structure in the game is still fundamentally about beating levels instead of

Figure 3: Screenshot of *Chicory* (left) and *Passpartout 2* (right). *Chicory* screenshot shows a fully colored level and the navigation mushrooms putting paints on the level when used. *Passpartout 2* shows the painting interface and a specific task for designing a car.

allowing oneself to engage in visual creativity directly. Games such as *Mario Party Flinger Painting* [43, 54] or *Splatoon* [16] also shift the incentive structure so that putting on paints is more about maximizing the area of coverage to win against the opponents. While one can extract these creations and find aesthetic value in them, this act of appreciation is not explicitly coded in the game's rules nor implicitly encouraged by the framing.

Both *Chicory* and *Passpartout 2* are painting/drawing games that advertise themselves as offering a game space for players to express their visual creativity. Both games narrativize the player's painting in a fictional setting. While *Passpartout 2* places the player character as a lost painter, *Chicory* builds its world around painting by mythologizing painters as wielders of *the Brush*, passed down from a line of wielders not dissimilar to royalty. In *Passpartout 2*, the player has to start from being homeless to ascend to being a celebrated painter in town, while in *Chicory*, the player character wields the Brush to fight back the mysterious corruption appearing in the world to bring back its colors. The main difference between *Chicory* and *Passpartout 2* is how they integrate painting into their game loops. *Chicory* is mainly a narrative-driven top-down 2D puzzle game, where the player navigates through environmental puzzles. The game has a coloring book aesthetic where the world is initially black and white. The player wields the brush to not only draw on the environment, but also to interact with in-game objects to navigate through the environment. On the other hand, *Passpartout 2*'s painting and the gameworld are rather separate interactions. The player navigates through the world, talks to characters, and buys painting tools. The painting starts only when the player puts down the canvas and starts to paint, with an interface not dissimilar to a very primitive drawing tool. Both games are shown in Figure 3. We'll discuss their designs regarding how they structure the player's creative process below.

4.1.1 Narrative Contextualization. The most direct complement that games can offer to painting is to provide additional context for the activity. Contextualization has a softer effect on manipulating the possibility space, as in it does not restrict possibilities on a software level, but rather makes certain decisions more likely. For example, painting existing characters in the game would be more likely to happen than not, as well as painting something tonally congruent with the game's setting. It also gives players the opportunity to role-play by asking questions such as: what would my character paint? for whom are they painting? what would be an appropriate painting for this narrative moment?

Outside of storytelling, there can also be unique contexts for one-off tasks that drastically change how the player engages with painting. For example, *Passpartout 2* has requests such as designing car paint jobs and theater props. Instead of a flat canvas, the player is given a unique 3D shape to paint over, as shown in Figure 3(right). In *Chicory*, the player character is also given similar tasks such as designing a shirt or the color of the displayed donut in a donut shop. These one-off tasks tend to have very visible consequences, such as the player seeing their designed car being driven by NPCs or being able to put on the shirt of their design. These unique painting opportunities can be read as alternatives to painting on a flat digital canvas. In a way, a flat canvas itself also constrains certain approaches to visual creativity. A flat red shape on a t-shirt or a car is a completely different decision than putting the same shape on a flat canvas (in the corner or at the center). Through this comparison, one can see that how players navigate through creative spaces is already conditioned by the site of engagement.

A flat and rectangular digital canvas already biases the space in certain ways just as much as designing for a shirt or a car.

4.1.2 Limited Choices. Unlike painting tools, painting games have no obligation to be feature-complete. The game can make expected painting features completely unavailable to shape the way players approach painting. In *Chicory*, the player only has access to four colors per in-game area, with the addition of white to start with. This incredibly limited palette effectively eliminates many possible decisions or approaches such as complex shading or values. In exchange, the players might more quickly start putting down colors on the canvas and experimenting with how different brush strokes, shapes or fillings affect the overall image. In addition, the majority of the colors available tend to be highly saturated and unnatural. This way, the player is unable to fill in an object with a natural color that correlates with its real-world appearance (grass → green). These design decisions point to an approach to coloring that resembles children's coloring books and the carefree experimentation that one can have without concern for "correctness."

This also speaks to the distinction between designing around *seeing for visual communication* and *seeing for recognition*. Color is a visual parameter that can itself evoke ideas and intentions without merely symbolizing objects or abstract ideas. From this point of view, *Chicory*'s deliberate limits on color also liberate color as a medium of expression in itself, while for text-to-image models, colors are subsumed for the function of recognition of objects and ideas.

Similarly, both *Chicory* and *Passpartout 2* exclude undo buttons, an almost fundamental feature of digital painting tools. While this absence is less jarring in *Chicory*'s casual and carefree approach to painting, it is rather missed in *Passpartout 2* — whose painting interface resembles that of painting tools, as seen in Figure 3 (right). Regardless of whether the undo button is intentionally excluded, it certainly encourages the player to paint past their mistakes, or to discover new possibilities within those mistakes. In the words of Bob Ross, it encourages treating mistakes as "happy accidents."

4.1.3 Shortcuts and Guides. It is common for tools to have user-friendly design features to make a particular type of use easier. For example, Photoshop's auto-align feature makes it easier to align different layers. These design choices are also observed in *Chicory* and *Passpartout 2*. For example, *Chicory* puts further emphasis on its coloring book design by making it easier to fill in colors with the brush. When the player holds down the draw button, a color circle will expand from the brush tip until it reaches the outline enclosing the shape, which is typically the edge of the ground, a cliff, or a house. *Chicory* does a particular interpretation of coloring through this animated "fill" feature: rather than emphasizing the manual act of filling out colors with a limited-sized brush through tight strokes, coloring can also be simply about choosing a single color for a shape and filling it in almost instantly. The hand is thus de-emphasized in this move, while the desire for the eye to see the result is accelerated. This is further supported by the fact that the canvas records colors in an absolute manner — each has to be one of the five colors and not anything in between. This design de-emphasizes pressure and repeated strokes of the hand, thus restricting its expressiveness (one can imagine pressure simulation or strokes with half transparency to allow subtler shaping of color values by the hand). In addition, if the player decides to fill in the color by brushing, the brush will softly stick to the outlines of the shape (the player can drag it further to unstick), which further promotes an approach to coloring that respects the lines on the image. These decisions do not technically change the possible outcomes, but they do manipulate the action space in such a way that color filling is much easier and more intuitive.

4.1.4 Loose and Opaque Evaluations. There is a deep tension between the ludic structure of games and creative activities. Lobanov, the creator of *Chicory*, argues in his GDC talk that creative activities should be separated from logical problem-solving and game progressions [36]. Otherwise, the player might emphasize the extrinsic motivations to move through the story and to obtain rewards rather than the intrinsic motivation found in creative acts. Rather, only observations should be provided as a way to react to the player's input (commenting on what's in the image without hard evaluation). There is no mechanism in the *Chicory* that forces the player to satisfy certain criteria in order to progress.

However, *Passpartout 2* implements a more subtle approach that still involves some form of algorithmic evaluation: through the money reward. While for the main story progression, there is no evaluation — the player can submit anything to pass the barrier — the game also allows the player to sell their artwork anywhere to anyone. The player can put a finished artwork on the table anywhere in the game, and nearby NPCs will walk to the table and offer money for the artwork. Different NPCs have different tastes for art and offer different amounts of money. But the underlying logic for each is completely obscured. The player can still learn to maximize certain features of their paintings for maximum earnings, but much is guesswork. *Passpartout 2* narrativizes this evaluation somewhat in two ways, 1) the act of figuring out the NPC's tastes reflects on the reality of being an artist professionally and 2) evaluation is attributed to specific NPCs, so any arbitrariness is not framed as an intrinsic design flaw. In addition, earning money is a purely positive reward that differs by the amount earned, which can lessen the player's desire to limit their creativity.

This is not to say that *Passpartout 2* puts more constraints on player expressions than *Chicory* because of the presence of evaluations. In fact, *Chicory* heavily restricts the player's exploration of visual creativity through its limited color palette choice and lack of stroke simulation (*Passpartout 2* simulates the dynamics of different brushes on the canvas), as well as the fact that the majority of coloring happens over the black-and-white outline of the level. *Passpartout 2*'s loose evaluation ("use a lot of blue") and opaque evaluation are used more to ground the player's creative decisions in the game world, and as the beginning of creative exploration, rather than as the end goal (instrumental play). Similar to narrative contextualization, it softly orients the player such that certain creative decisions are more likely than others ("use blue" → what ideas are possible by simply painting with blue?).

4.2 Design Features of Experimental Text-to-Image Games

Here we survey three games that use text-to-image models as their main interaction: *Bureau of Multiverse Arbitration* (*BoMA*), *Keyo Against Humanity*, and *Laughprop*. These games also promote visual

creativity in that the games encourage the players to generate novel results using text-to-image models. Because of the newness of text-to-image models and their high hardware requirement (mostly GPU memory), most games exist still in experimental, demo, and student project spaces. Some games are discontinued without proper historical documentation and released artifacts. Specifically, there are two games that we surveyed with no currently playable version available: *Keyo Against Humanity* is a student project developed in an AI incubator program largely aimed to train software engineering skills with only presentation video available. *BoMA* is a game run on Discord with timed tasks and feedback provided in the form of YouTube videos. The game ended around December 2022 and is no longer maintained. But the chat history remains preserved on the Discord server. We studied the chat history to infer the relevant design features and play experience involved. Although many artifacts are lost to history, it is nevertheless interesting to see attempts to gamify text-to-image interactions, in a way that reinforces our thesis regarding the distinction between seeing for visual communication and seeing for recognition.

All three games follow a similar high-level structure: use text-to-image models to create novel images, and the players will vote on the winner — with the exception of *Keyo Against Humanity*, which asks the player to then write captions on the generated image so the players can vote on the winning caption. Only *BoMA* contextualizes the player in a fictional universe, as an agent of the Bureau of Multiverse Arbitration, which offers "multiverse consultation" to clients with decision paralysis. To put it simply, the bureau uses the multiverse search engine (text-to-image model) to offer solutions to a client's problems. In addition to a standard generating-voting game mode, *Laughprop* offers two different modes for the player to engage with text-to-image, which are discussed below.

4.2.1 Narrative Contextualization and Game Premises. Both *BoMA* and *Laughprop* offer prompts that the players must consider for their text-to-image creation. In *BoMA*, these are framed as the client's requests. The prompt is typically provided in a short one-line request, for example "Submit a **human/plant** hybrid for a megaflora greenhouse utopia" or "Submit a sophisticated **animal mashup** for an interstellar dog show." The prompts are obvious techniques to constrain and kickstart the creativity of the player. The world-building of a multiverse arbitration should not be ignored here. The outlandishness of the context of multiverse arbitration, combined with self-aware humor in many of the voice deliveries, provides a safe space for bizarre combinations of objects and ideas while still remaining within the bounds of the fiction. The many inaccuracies and distortion of the images generated can also be overlooked without directly breaking the diegesis.

BoMA's presentation is in direct contrast to *Laughprop*, which is without fictional context. The two text-to-image game modes in *Laughprop* also put soft constraints on the input of the players. "It's a mood" mode simply gives players the same theme/prompt such as "a hairy situation" or "best place to hide in a zombie apocalypse" and asks them to describe scenes for text-to-image generation. "I'd watch that" asks the players to pick a movie and to cast different actors for the role. The player is constrained to inputting what would replace a specific character in a specific movie (e.g., Luke Skywalker in *Star Wars*). "I'd watch that" is a rather direct and explicit constraining to recognizable objects in many ways. This shows that, although *Laughprop* in general does not have explicit fictional contextualization like *BoMA*, it does not equate to the player having the maximum freedom to make any creation as they desire. Rather, the player might default to their habituated way of being imaginative and creative. A player playing "I'd watch that" might default their immediate choices to existing actors or characters, or maximally bizarre creatures.

4.2.2 Curation-based Play. Given the large volume of images generated with text-to-image models, the role of the player veers towards curation rather than direct manipulation of images. Many games encourage comparing multiple image outputs before committing to a submission. *Laughprop* would generate four images for the player to pick from, while *Keyo Against Humanity* also has a "submit" button that allows the player to re-generate their image or re-work their prompt until they are satisfied with the result. In *BoMA*, the player can keep generating images until they find one that warrants clicking on the "submit solution" button.

Curation play is particularly intense in *BoMA*. Curating a large volume of images without direct manipulation encourages a very particular type of player behavior. For example, because each generation has different seeds, the player might simply re-enter the same prompt until a satisfactory image shows. The chat history quickly accumulates as the same player constantly re-generates their images in a short amount of time. The prompting approach is also specific to the characteristics of the stable diffusion model. The words used in the prompt are split into two categories: one that is supposed to be recognized such as "white poodle" and "yellow flight suit," while the others are simply put in to deviate from the default aesthetic of the generator to improve on a vague sense of quality, for example: "cinematic, color grading, shot on 50mm lens, ultra-wide angle, depth of field, hyper-detailed, beautifully color-coded, insane details, intricate details, soft lighting, volumetric, global illumination, chromatic aberration." The latter use of words, while still injecting ideas to manipulate the image, isn't strictly for purposes of recognition. Rather, it is better conceptualized as simply a technique to manipulate the image that is very specific to the text-to-image model used (stable diffusion).

The process of curation in these games is perhaps when the player engages with seeing for visual communication the most. Players must engage with the image directly for its aesthetic effects, barring other questions of recognition, such as how well the human anatomy is portrayed. But these aesthetic engagements are vague, un-directed, and can vary in sophistication from person to person. Oftentimes, the aesthetics are boiled down to a vague sense of quality that can vary greatly depending on the context. The interface also does not afford much control when problems are identified through more fundamental visual units (the color is not quite right, the highlight is too much, or the composition is misaligned), thus reducing the usefulness of seeing as visual communication.

4.2.3 Social Evaluation. All three games utilize voting to determine the winner. While *Laughprop* simply presents the winner image on the final screen, *BoMA* provided more flourishes through final observation reports, each a video with the field agent's voice narrating over the winning images; see [2] for an example. Oftentimes, the video simply mentions the most notable feature (typically the

objects or ideas involved) of the winning images. On the other hand, *Keyo Against Humanity* asks the players to write captions for the generated image; the players then vote on which caption wins the game.

Social evaluation is a direct and simple-to-implement mechanic to introduce a win/lose state into a game. But how to do it well is a rather tricky design question — because fostering appropriate social space and mood is the key to allowing the sociality to enhance the experience, rather than diverting it. *Keyo Against Humanity* especially lacks consideration in this area. The social evaluation can develop a dynamic where the answers are more crowd-pleasers than anything related to the image generated. *Laughprop* has a very short game loop in which simple prompting → curating image → voting for a winner can end in a matter of minutes. It becomes difficult to take the game seriously in that it takes more effort to set up a game than to explore the potential of image generation. *BoMA* manages this social space well by having a large number of internet users in a Discord server with channels for casual chatting and game interactions. The voting phase is also a timed event to create a certain sense of anticipation and camaraderie among the players. As a result, the players appear to take the act of voting more seriously — in that it is more about evaluating the images, rather than any interaction unrelated to the game.

5 IDEATING CRITICAL GAME ELEMENTS

In Section 3, we propose the distinction of *seeing as visual communication* and *seeing as recognition* to talk about how text-to-image interfaces and painting interfaces differently habituate the user's orientation to their visual creativity. These different types of habituation, or two different ways of seeing, are the theoretical foundations of the kind of critical literacy that we wish to communicate through video games. From this perspective, we see our two different ways of seeing as close to strong concepts in HCI [28], in that this not have the scale of a full theory, but is abstract enough to connect user/player behavior and design approaches, as well as being potentially generative of new design elements.

This section seeks to combine both the theory work in Section 3 and the light survey work in Section 4 to take the initial step towards developing games interested in interrogating text-to-image AI models. The format will be proposing potential game design elements that can make such a theoretical distinction explicit.

It is worth expanding on what it means to "make such distinction explicit." It means that the design elements are mostly going to be motivated by the two different ways of seeing — the concepts becoming generative of design elements. The two ways of seeing have rather complex relations to each other and are in no way obvious on the surface. Indeed, when appreciating visual art, one cannot disentangle the recognized concept or idea and the artistic ways the visual elements are manipulated. Although our theoretical contribution argues that two types of interfaces habituate two different ways of seeing, the power of this distinction cannot be separated from the context around which two ways of seeing become valued, acknowledged and circulated. While generating one image from a text-to-image model might not immediately habituate the user to seeing as recognition, the habituation becomes more distinct when the user is embedded within the text-to-image community or committed to create images for financial purposes. It is within the context of use, especially through player-AI interaction [64], that the two different ways of seeing become distinct and critically useful. The implementation of design elements and game mechanics is thus understood as shaping of that context in a ludic manner. In the Mechanics-Dynamics-Aesthetics (MDA) model [29], one can understand the proposed design elements here as mechanics to give rise to the two ways of seeing as gameplay dynamics.

But it is not enough to simply design for gameplay dynamics. The ultimate goal is a game that can develop critical literacy within the player. (To reiterate, critical literacy here means that meta-level knowledge about the game, both cognitive and linguistic, can be used to critique discourses around AI outside the game [24].) Thus, the gameplay dynamics are not just interesting interactive ideas that the game explores. Rather, the game dynamics have to *point to* or *say something about* certain things in the world. Here we take inspiration from Ian Bogost's concept of procedural rhetoric to understand the design of game structures as persuasive expression — that the inclusions and exclusions of entities and their interactions are expressing certain perspectives about the world [8]. This is distinct from merely using ludic structures around educational content (such as game-like quizzes) to develop literacy. We follow Bogost's definition of "procedural literacy." The game design itself is understood as offering insights into "particular worlds and particular relationships" that drive certain experiences through "interacting with the abstract models of specific real or imagined processes" [8, p241-260]. What this means for our design is that *any inclusion of design elements cannot be arbitrary, but must communicate certain perspectives about text-to-image models and visual creativity outside the game itself.*

The baseline premise of the game is to offer both text-to-image and traditional painting interfaces for the player to work towards a goal. This goal can simply be a given prompt such as "draw something sad" or "a person feeling sad," or it can employ obscure evaluations that the player needs to discover through trial and error. The main aim is to allow the player to develop a certain familiarity with visual creation. By this we mean literacy within the visual creative domain through embodied acquisition (skills and knowledge developed through hands-on painting) [24, 25]. Other design elements will be motivated towards meta-level criticality (guiding how the player thinks about their act of painting). The following proposed design elements include both considerations: embodied acquisition of visual creativity and meta-level criticality. It should be noted that these proposals and effects are speculative, and they all require extensive prototyping and playtesting for validation and refinement.

5.0.1 Limited Choices and Gamification. To reiterate the introduction, we target the general audience who has some general interest in the issues around AI. As a result, we cannot assume that they are fluent with visual creativity. Our light survey of painting games indicates that limited choices and software-based assistance can help the player get into the creative mood more quickly and easily. This can mean simply asking the player to commit to a certain color palette beforehand, similar to *Chicory*, or to limit themselves to certain brushes. The game can also simulate a certain dynamic of blending (e.g., oil painting or watercolor) to not only make each

brush stroke more dynamic, but also to commit the player to a certain style of painting. These features encourage the player to experiment with what's available to them, rather than to stay in choice paralysis.

In short, the limitation/assistance makes the act of painting more experiment-based and, hopefully, fun. The same principle can be applied to text-to-image interfaces, although completely different techniques are deployed. One idea is to utilize different natural language techniques to add random words and phrases into the prompt, or simply scramble the player's prompt (such as shuffling its grammar structures). One simple technique would be adding random words with Markov generations trained on public domain literature. The central goal is to allow the player to be playful and improvisational with text, words and concepts, so it's easier to gain creative momentum.

5.0.2 Changable Relationships between Interfaces. Inspired by Augusto Boal's theater of the oppressed, Gonzalo Frasca proposes the concept of meta-simulation to think about games that help develop criticality around a simulated domain [21]. Essentially, the player can change the parameters and rules of the simulation to see how the simulation produces different results. This prompts the player to imagine alternatives to existing configurations and thus to develop criticality over them if they have obvious real-world referents. This is an explicit gamification on the meta-level of a domain, because it explicitly prompts the player to not take for granted the rules around a particular domain. Similarly, one possible design for the critical game is to allow the player to choose between using a painting interface, or a text-to-image interface, or some blend between the two.

One possible player takeaway could be about the affordances and limitations of each tool: What is each good for, given particular aesthetic goals? The game can provide a visualization of which tool contributes to which part in the image. A possible hypothesis based on this design is that the player would recognize they are more playful with colors and shapes than simply using text-to-image generation. Another takeaway might be shaped by adding constraints that motivate the use of one interface over the other. For example, time constraints would push towards more use of text-to-image models. Costs over the two interfaces (such as computational cost or man-hour cost) can also push players towards one over the other. Given enough contextualization, the consideration of cost here can prompt the player to become critical of the circumstances in which cost becomes important, and how these considerations can determine the adaptation of one interface over the other, and thus possible creative results.

5.0.3 Curation and Commitment. Our light survey of text-to-image games showed a pattern of use that centers curation of images. This is because text-to-image prompting has a rather short iterative loop that returns dramatically different results. Random seeds also encourage the user to keep generating images without much intentional authoring. There is an opportunity here to make the user compare how they treat the text-to-image results differently when they are limited by how many requests they can send to the model. Especially when the text-to-image model has to be used in combination with the painting interface, it may force the player to commit to certain aspects of the image that they would not otherwise. One hypothesis is that the player would be more habituated to seeing as recognition, while having to commit to a generated image pushes them to seeing as visual communication through closer examination of the visual content.

5.0.4 Narrative Contextualization. We see narrative contextualization as an important part of developing criticality. This is mainly for two reasons. First, narrative contextualization is key to connecting gameplay elements to the real-world phenomena being referenced. (On this point, we disagree with the naive proceduralist view that overly emphasizes interactive dynamics as rhetorical acts [8, p241-260].) Indeed, [4] has shown that abstract games tend to communicate their central messages less effectively. For example, the introduction of limitations and costs can be framed as cost-cutting in the corporate setting. So the mechanics are not perceived as included purely for gameplay reasons, but rather are recognized as interpretations of how corporate settings shape the dynamics of creative acts.

Second, contextualization can also introduce proper value assertions and elicit the intended affective responses. Flanagan & Nissenbaum mention the game *Layoff* [56] that had to add additional information about the people being laid off in order to induce the feeling of guilt in the player when they interact with the core mechanics to fire people [20, p133]. Since the issues of text-to-image (and AI in general) inevitably involve the larger concerns of labor precarity, we see it as important to communicate our values in these areas. It is important to not portray people as simply means for particular ludic goals.

5.0.5 Social Justification and Meaning-making. Since we define visual creativity as how the author communicates their intentions through visual creations, one way to make the player realize how different interfaces shape their intentions is to make the communication of those intentions part of the gameplay. One possible design is to introduce social aspects into the game, such as the player having to argue for their final image (e.g., why certain decisions were made) either to the host or to other players. Some inspiration can be taken from the social evaluation mechanics found in text-to-image games — such as audience and player voting and ranking. This could be a particularly intriguing feature for further development of our theories, as how the players communicate their intentions could be studied to validate and complicate our theory in Section 3.

In order to assist the player in forming their artistic intentions, subtle suggestions can be provided to assist meaning-making. For example, poetic languages can pop up when picking colors such as blue — "color of the sea, of the sky, of melancholy." A more elaborate feature could be to use large language models to assist the player in making decisions, interpreting what's on the canvas and providing new ideas.

6 CONCLUSION

Overall, this paper's main goal is to offer a humanistic framework around text-to-image and to explore how such ideas can be communicated through games. Specifically, we wish to illustrate how producing aesthetic images using text-to-image is far from a replacement for traditional drawing and painting. Rather, it is a different

way of creating and doing that constructs different relationships between the creator and the output. We propose the difference between *seeing for visual communication* and *seeing for recognition* as the core difference between the two interfaces. We also survey games with these two interfaces to identify design features that can assist the players in becoming visually creative. We propose possible game design elements to illustrate how we can communicate this critical literacy through the medium of video games. Needless to say, much prototyping, development, playtesting, and iteration have to take place in the future.

REFERENCES

[1] Aconite. 2022. Bureau of Multiversal Arbitration. https://www.bureauofmultiversalarbitration.com/

[2] Aconite Aconite. 2022. Final observation report // Wolfina's den. https://www.youtube.com/watch?v=--bkUmpXSQ0

[3] Teresa M Amabile, Regina Conti, Heather Coon, Jeffrey Lazenby, and Michael Herron. 1996. Assessing the work environment for creativity. *Academy of management journal* 39, 5 (1996), 1154–1184.

[4] Barrett R Anderson, Christopher R Karzmark, and Noah Wardrip-Fruin. 2019. The psychological reality of procedural rhetoric. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*. 1–9.

[5] Sofian Audry. 2021. *Art in the age of machine learning*. Mit Press.

[6] Margaret A Boden. 2004. *The creative mind: Myths and mechanisms*. Psychology Press.

[7] Margaret A. Boden and Ernest A. Edmonds. 2009. What is generative art? *Digital Creativity* 20, 1-2 (2009), 21–46. https://doi.org/10.1080/14626260902867915 arXiv:https://doi.org/10.1080/14626260902867915

[8] Ian Bogost. 2010. *Persuasive games: The expressive power of videogames*. mit Press.

[9] AI Camp. 2022. Stable Diffusion Multiplayer Game - Keyo Against Humanity Launch. https://www.youtube.com/watch?v=CSGj1yUllpc

[10] Eva Cetinic and James She. 2022. Understanding and creating art with AI: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 18, 2 (2022), 1–22.

[11] Jenova Chen. 2007. Flow in games (and everything else). *Commun. ACM* 50, 4 (2007), 31–34.

[12] Kate Compton and Michael Mateas. 2015. Casual Creators.. In *ICCC*. 228–235.

[13] Chris Crawford. 2003. *Chris Crawford on game design*. New Riders.

[14] Mihaly Csikszentmihalyi. 1997. Flow and the psychology of discovery and invention. *HarperPerennial, New York* 39 (1997), 1–16.

[15] Gilles Deleuze and Francis Bacon. 2003. *Francis Bacon: The logic of sensation*. U of Minnesota Press.

[16] Nintendo Entertainment Analysis I& Development. 2015. Splatoon.

[17] Baker Diane. 2018. Drawing excercise instructions based on: Drawing on the right side of the brain. - ppt download. https://slideplayer.com/slide/9222810/

[18] Kayleigh Donaldson. 2023. What ai (and everyone else) misunderstands about Wes Anderson. https://www.pastemagazine.com/movies/wes-anderson/wes-anderson-ai-aesthetic-misunderstand

[19] Betty Edwards. 1997. Drawing on the Right Side of the Brain. In *CHI'97 Extended Abstracts on Human Factors in Computing Systems*. 188–189.

[20] Mary Flanagan and Helen Nissenbaum. 2014. *Values at play in digital games*. MIT Press.

[21] Gonzalo Frasca. 2001. *Videogames of the oppressed: Videogames as a means for critical thinking and debate*. Master's thesis. School of Literature, communication, and culture, Georgia Institute of ....

[22] 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*. 1–18.

[23] Flamebait Games. 2023. Passpartout 2: The Lost Artist. https://store.steampowered.com/app/1571100/Passpartout_2_The_Lost_Artist/

[24] James Paul Gee. 1989. What is literacy. *Journal of education* 171, 1 (1989), 18–25.

[25] James Paul Gee. 2003. What video games have to teach us about learning and literacy. *Computers in entertainment (CIE)* 1, 1 (2003), 20–20.

[26] James Paul Gee. 2005. Learning by design: Good video games as learning machines. *E-learning and Digital Media* 2, 1 (2005), 5–16.

[27] Adam Gopnik. 2023. What can A.I. Art teach us about the real thing? https://www.newyorker.com/culture/cultural-comment/what-can-ai-art-teach-us-about-the-real-thing

[28] Kristina Höök and Jonas Löwgren. 2012. Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Transactions on Computer-Human Interaction (TOCHI)* 19, 3 (2012), 1–18.

[29] Robin Hunicke, Marc LeBlanc, Robert Zubek, et al. 2004. MDA: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI*, Vol. 4. San Jose, CA, 1722.

[30] Wilawan Inchamnan, Peta Wyeth, and Daniel Johnson. 2013. Does activity in computer game play have an impact on creative behaviour?. In *2013 IEEE International Games Innovation Conference (IGIC)*. IEEE, 77–84.

[31] Rebecca Jennings. 2023. Ai Art Freaks Me Out. so I tried to make some. https://www.vox.com/culture/23678708/ai-art-balenciaga-harry-potter-midjourney-eleven-labs

[32] Mukund Kapoor. 2023. Negative prompts in stable diffusion: A beginner's guide. https://www.greataiprompts.com/imageprompt/what-is-negative-prompt-in-stable-diffusion/#google_vignette

[33] James C Kaufman and Ronald A Beghetto. 2009. Beyond big and little: The four c model of creativity. *Review of general psychology* 13, 1 (2009), 1–12.

[34] Jingyi Li, Eric Rawn, Jacob Ritchie, Jasper Tran O'Leary, and Sean Follmer. 2023. Beyond the Artifact: Power as a Lens for Creativity Support Tools. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[35] Greg Lobanov. 2021. Chicory: A Colorful Tale. https://store.steampowered.com/app/1123450/Chicory_A_Colorful_Tale/

[36] Greg Lobanov. 2022. Gamifying creativity in "chicory: A colorful tale". https://www.youtube.com/watch?v=UTCB9oqWidQ

[37] Gary Marcus, Evelina Leivada, and Elliot Murphy. 2023. A Sentence is Worth a Thousand Pictures: Can Large Language Models Understand Human Language? *arXiv preprint arXiv:2308.00109* (2023).

[38] Scott McCloud. 2002. *Reinventing comics: How imagination and technology are revolutionizing an art form*. HarperCollins Publishers.

[39] Goose Minded. 2022. SuchArt: Genius Artist Simulator. https://store.steampowered.com/app/1293180/SuchArt_Genius_Artist_Simulator/

[40] Piet Mondrian. 2019. Composition with Red, Blue and Yellow. https://collection.kunsthaus.ch/en/collection/item/2455/

[41] Claude Monet. 1894. Rouen Cathedral, West Façade. https://www.nga.gov/collection/art-object-page.46524.html

[42] nat04194 nat04194. 2022. How to identify AI generated images.A thread pic.twitter.com/nrnblvjzdv. https://twitter.com/nat04194/status/1599928975908016129?lang=en

[43] NDcube. 2012. Mario Party 9.

[44] C Thi Nguyen. 2020. *Games: Agency as art*. Oxford University Press, USA.

[45] Jonas Oppenlaender. 2022. The creativity of text-to-image generation. In *Proceedings of the 25th International Academic Mindtrek Conference*. 192–202.

[46] Jackson Pollock. 2018. Number 1A, 1948. https://www.moma.org/collection/works/78699

[47] Seyedahmad Rahimi and VJ Shute. 2021. The effects of video games on creativity: A systematic review. *Handbook of lifespan development of creativity* 37 (2021).

[48] Seyedahmad Rahimi and Valerie J Shute. 2021. First inspire, then instruct to improve students' creativity. *Computers & Education* 174 (2021), 104312.

[49] R Keith Sawyer. 2011. *Explaining creativity: The science of human innovation*. Oxford university press.

[50] Michael Sellers. 2017. *Advanced game design: a systems approach*. Addison-Wesley Professional.

[51] Ben Shneiderman. 2000. Creating creativity: user interfaces for supporting innovation. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 1 (2000), 114–138.

[52] Ben Shneiderman. 2002. Creativity support tools. *Commun. ACM* 45, 10 (2002), 116–120.

[53] Valerie J Shute and Fengfeng Ke. 2012. Games, learning, and assessment. In *Assessment in game-based learning: Foundations, innovations, and perspectives*. Springer, 43–58.

[54] ShyGuyStation. 2019. *Mario Party 9 Minigames - Flinger Painting*. https://www.youtube.com/watch?v=94xVObvhxCI

[55] Christian Sivertsen, Guido Salimbeni, Anders Sundnes Løvlie, Steve Benford, and Jichen Zhu. 2024. Machine Learning Processes as Sources of Ambiguity: Insights from AI Art. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*.

[56] Tiltfactor. 2009. Layoff.

[57] Bart Trzynadlowski and Steph Ng. 2023. Laughprop: Stable Diffusion Party Game. https://github.com/trzy/Laughprop

[58] Thomas Van Den Berge. 2021. Vermillion - VR Painting. https://store.steampowered.com/app/1608400/Vermillion__VR_Painting/

[59] Graham Wallas. 1926. *The art of thought*. Vol. 10. Harcourt, Brace.

[60] Ota Yoshitaka. 2013. What is the 'Haptic': Consideration of Logique de la Sensation and Deleuze's theory of sensation. *The Japanese Society for Aesthetics* 17 (2013), 13–24.

[61] ZackBellGames. 2015. INK. https://store.steampowered.com/app/385710/INK/

[62] Robert Zeller. 2016. The Figurative Artist's Handbook: a contemporary guide to figure drawing painting and composition.

[63] Liang Zeng, Robert W Proctor, and Gavriel Salvendy. 2011. Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity?

Creativity Research Journal 23, 1 (2011), 24–37.

[64] Jichen Zhu, Jennifer Villareale, Nithesh Javvaji, Sebastian Risi, Mathias Löwe, Rush Weigelt, and Casper Harteveld. 2021. Player-AI interaction: What neural network games reveal about AI as play. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[65] Robert Zubek. 2020. *Elements of game design.* MIT Press.

[66] Joanna Zylinska. 2020. *AI art: machine visions and warped dreams.* Open Humanities Press.