Who did it? How User Agency is influenced by Visual Properties of Generated Images

Johanna K. Didion
School of Computer Science,
University College Dublin,
Dublin, Ireland
Sensorimotor Interaction,
Max Planck Institute for Informatics,
Saarland Informatics Campus
Saarbrücken, Germany
johanna.didion@ucdconnect.ie

Krzysztof Wolski
Max Planck Institute for Informatics,
Saarland Informatics Campus
Saarbrücken, Germany
kwolski@mpi-inf.mpg.de

Dennis Wittchen
University of Applied Sciences
Dresden, Germany
Sensorimotor Interaction,
Max Planck Institute for Informatics,
Saarland Informatics Campus
Saarbrücken, Germany
dennis.wittchen@htw-dresden.de

David Coyle
School of Computer Science,
University College Dublin,
Dublin, Ireland
d.coyle@ucd.ie

Thomas Leimkühler
Max Planck Institute for Informatics,
Saarland Informatics Campus
Saarbrücken, Germany
thomas.leimkuehler@mpi-inf.mpg.de

Paul Strohmeier
Sensorimotor Interaction,
Max Planck Institute for Informatics,
Saarland Informatics Campus
Saarbrücken, Germany
paul.strohmeier@mpi-inf.mpg.de

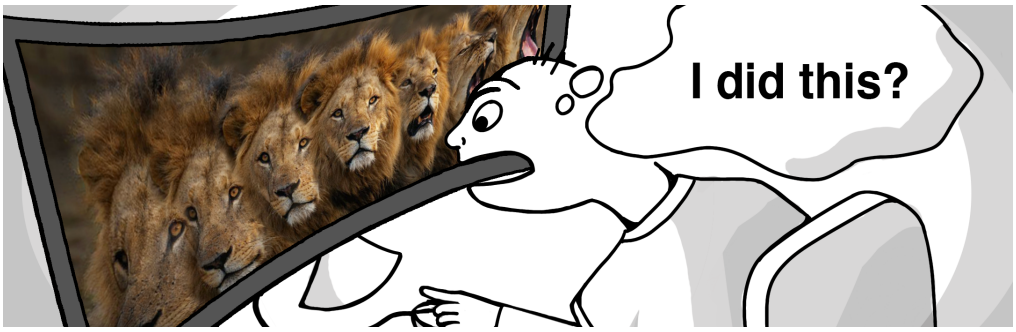

Figure 1: The current study investigates the experience of agency when creating images using generative AI. Resulting images range from uninteresting to awe-inspiring to jarring.

ABSTRACT

The increasing proliferation of AI and GenAI requires new interfaces tailored to how their specific affordances and human requirements meet. As GenAI is capable of taking over tasks from users on an unprecedented scale, designing the experience of agency – if and how users experience control over the process and attribution of the outcome – is crucial. As an initial step towards design guidelines for shaping agency, we present a study that explores how properties of AI-generated images influence users' experience of agency. We use two measures; temporal binding to implicitly estimate pre-reflective agency and magnitude estimation to assess user judgments of agency. We observe that abstract images lead to more temporal binding than images with semantic meaning. In contrast, the closer an image aligns with what a user might expect, the higher the agency judgment. When comparing the experiment results with objective metrics of image differences, we find that temporal binding results correlate with semantic differences, while agency judgments are better explained by local differences between images. This work contributes towards a future where agency is considered an important design dimension for GenAI interfaces.

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-0628-8/24/10
https://doi.org/10.1145/3654777.3676335

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; *HCI theory, concepts and models*; • **Applied computing** → Psychology; • **Computing methodologies** → Machine learning.

KEYWORDS

agency, generative AI, image generation, user experience

ACM Reference Format:
Johanna K. Didion, Krzysztof Wolski, Dennis Wittchen, David Coyle, Thomas Leimkühler, and Paul Strohmeier. 2024. Who did it? How User Agency is influenced by Visual Properties of Generated Images. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3654777.3676335

1 INTRODUCTION

Generative artificial intelligence (GenAI) has been hailed for its potential to fundamentally transform human society [1, 17, 101] and has become a topic of high interest within the HCI community [18, 81, 84]. One area of discourse is the intersection between GenAI and creative practices like image or text generation [73]. The impact of GenAI on creative fields has sparked intense debate, ranging from criticism for its potential to replace human artists and writers, to praise for its ability to democratize content creation [29].

This discourse around GenAI predominantly focuses on *technical* and *legal* considerations, such as the scale and sources of training data and the legal implications of generated content. These discussions often ignore the experiential dimension of interacting with GenAI, specifically, how users *subjectively experience* agency – control over the GenAI process and attribution of the outcome. There is a complex interplay between technical, legal and experiential properties of interacting with AI, which are all crucial yet distinct dimensions to consider.

Moreover, the relationship between user agency and the effectiveness of the AI algorithm presents inherent tension. While generating novel, unexpected solutions is a valued capability of GenAI, it inherently requires users to relinquish detailed control over the process. Even though human control is a desirable design goal in human-computer interaction (HCI), one might reasonably argue that good GenAI necessarily reduces users' control.

This suggests that for the interaction with GenAI, special attention should be given to how the interface design deals with users' experience of agency. For example, it might be desirable for an application to provide a heightened sense of agency for users to experience a creative flow, while for another application, it may be desirable for users to feel less agency when the AI is active, to prevent a confusion of attribution. The experience of agency then becomes a design material to create desirable GenAI interfaces.

This paper explores this intersection of the experience of agency and GenAI in the context of image generation. Through an experimental platform based on DragGAN [68], we examine how deviations from user expectations in AI-generated images affect perceived agency. Our study contrasts two scenarios: a *realistic* condition where the AI alters the positioning of human and animal faces in images as directed by the user, and a *changed* condition where the AI additionally transforms these faces – for example, a lion might become a fox – thereby introducing novel content not requested by the user into the output. We supplement the two conditions with two control conditions in which *distorted* and completely *abstract* images are generated, respectively. We measure the difference in experienced agency for all conditions using both implicit measures to capture the pre-reflective sense of agency and explicit measures to capture the judgment of agency.

We also analyze the differences between images from all four conditions using four objective metrics (l1, Histograms, LPIPS, DINO) and compare these measures to the experimental results. The resulting data is compatible with the hypothesis that user judgments are influenced by changes in local features of the images, and the hypothesis that pre-reflective sense is based on large semantic differences in images.

Our investigation into the subjective user experience of agency (Figure 1) when interacting with GenAI contributes to the broader discourse on how we, as a culture, wish to engage with artificial intelligence.

2 RELATED WORK

This section provides an overview over agency and a review of current research in the field. It also gives an introduction to GenAI and methods for shaping image generation.

2.1 Agency

The sense of agency, as discussed in psychology and cognitive science, is our experience of control of the world around us as we act in it [37]. This sensation operates subconsciously; interestingly, it is often most noticeable when absent. When we acknowledge our actions with statements like "I did that", it is not just about taking credit for our actions, but also about expressing belief in our ability to control and taking ownership over and responsibility for the actions in our everyday lives [21], all of these aspects of agency influence how we see and interact with the world.

Our sense of agency forms early on in life. Within the first 3-6 months, young infants learn that their movements can have an impact, e.g., shaking a rattle causes a noise. This encourages exploration, and research suggests that this in turn helps infants to develop greater control of movement and also a sense of agency [8]. Similarly, infants also experience agency through eye contact and episodes of shared attention with their parents. They respond to and mirror gaze direction and affective signals, fulfilling the children's need for self-efficacy [70]. The experience that one's actions have an effect on the world and on other people is foundational to one's ability to navigate most aspects of life [69].

Our sense of agency is malleable by context [65]. For example, when using an Ouija board, people often feel reduced control over the planchette, which can be an eerie experience [5]. Other activities, such as gambling, can create situations where people experience more control than they actually have [40]. Such over- and underestimates of agency can be explicitly designed. For example, in HCI, studies have demonstrated that the latency with which one receives feedback [10, 94] and the selection of input modality [11, 12, 21, 57] both have a strong effect on perceived agency.

The importance and fragility of the sense of agency indicate that the design of our direct environment and the technologies within it should provide and support the experience of agency. Research shows that different aspects of interactions influence the human sense of agency. Our work aims to identify such aspects in the context of interactions with GenAI – especially image generation.

2.1.1 Modelling and Measuring Agency. A commonly used framework for understanding agency is the comparator model. It suggests that the central nervous system functions as a control mechanism,

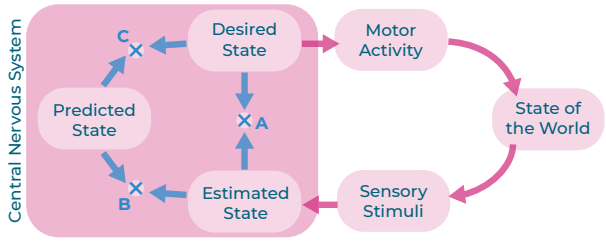

Figure 2: To know if we have achieved a goal, we must compare the desired state of the world with what we estimate the state of the world to be (comparator A). To understand if we ourselves achieved that goal or if other factors were involved, we can draw a comparison with a prediction of our actions (comparator B). Finally, comparing our prediction to our desired state (comparator C) gives us information about whether we can control the current system sufficiently to achieve our goals.

optimizing behaviour to achieve desired outcomes. Actions modify the world, and sensory feedback helps construct a model of this changed state. By comparing the desired and actual states (Figure 2, comparator A), adjustments can be made to reach an optimal state [32, 89].

Our sense of agency is assessed when these desired and estimated states are compared to an internal prediction of the consequences of our actions [31]. Such an internal prediction is necessary because with sensory information alone, one cannot distinguish between actions caused by oneself and those caused by external factors. To make such a distinction, one must compare the *predicted outcome* of one's action with the *estimated state* based on sensory information (Figure 2, comparator B). If there is no discrepancy between the states, then we can assume the state of the world was caused by ourselves, while differences between prediction and estimation are attributed to external forces. This comparison helps determine *self-attribution* of our actions [32, 89]. Finally, we must also consider comparing our desired state with the predicted state (Figure 2, comparator C). This comparison is often seen as relating to our sense of *control*: If even the prediction of the outcome of our actions does not match our desired state, then we lack the control to achieve this outcome.

An alternative theoretical position that downplays the specific contribution of the motor system is put forward by Wegner et al. [96–98]. In this view, the sense of agency is produced by a more general-purpose cognitive system that monitors the relationship between thoughts (i.e. intentions), actions and their outcomes, with the mind inferring and reconstructing a path between conscious intention and effect [96]. Here, the sense of agency is taken to be a reconstructive inference that one's intention has caused an external event. Our mind essentially tells itself a story, one about the causal link between our actions and their outcomes. If the story is plausible, we infer (experience) a sense of agency. If it is not plausible, our sense of agency is diminished. Wegner and Wheatley argue that plausibility requires three conditions: priority, consistency, and exclusivity [98].

In the present work, we assume a two-level model of agency: a high level conscious agency judgment, which is influenced by a low level pre-reflective sense of agency (Figure 3).

2.1.2 Agency Judgements. To determine if someone experiences agency, we might simply ask. This approach was used by Nishida et al. [67] when they investigated whether users might attribute a computer's action to themselves if the outcome matched their intention. According to self-reports, if the interval between the computer's action and their own intention was small, users would attribute a system's action to themselves. However, when the discrepancy between expected and actual outcomes grew too large, users reported a diminished sense of agency. This aligns with the predictions of comparator B (Figure 2) of the agency model discussed here.

In a subsequent study, Tajima et al. [90] examined a task involving a correct action and an incorrect one. Using the same methodology, they found that users tend to self-attribute correct outcomes and attribute incorrect ones to the system. This suggests that, much like the causal history-based model [98], additional factors, like cognitive dissonance reduction or the social desirability of outcomes, might influence one's judgment of agency upon reflection.

These findings indicate that the sense of agency, as it arises from interaction, and the judgment of agency, as reported upon reflection, do not necessarily align [89]. There seems to be a bottom-up process leading to a pre-reflective sense of agency and a top-down process resulting in a judgment of agency [89] (Figure 3). Some argue that only the top-down judgment of agency is relevant for HCI, as it is a conscious experience. However, the judgment of agency seems to build upon the sense of agency (e.g., [67]). Furthermore, it is the factors leading to the bottom-up sense of agency that are primarily under our control as interface designers.

The present study explores both the pre-reflective sense of agency as well as explicit judgements of agency. However, as we believe the judgements to be based on the sense of agency, our primary interest is to better understand the bottom-up, pre-reflective sense of agency.

2.1.3 Measuring Pre-Reflective Agency. If we care about the pre-reflective sense of agency, and we cannot ask people about it, how

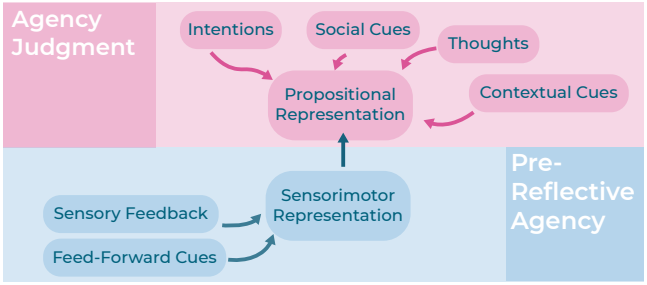

Figure 3: Sensory feedback (the information we infer from the world and feed-forward cues, that is, the modeled effects of our actions) provides a sensorimotor representation of agency (blue). Once we actively reflect on what has occurred, we consciously make a judgment of agency (pink). This judgment is informed by the pre-reflective sense of agency, but is interpreted and shaped by personal goals, intentions, thoughts, and social cues.

do we measure it then? There are two measurable perceptual phenomena often assumed to correlate with agency: *sensory attenuation* [53] and *temporal binding* [22]. *Sensory attenuation* is based on the observation that self-caused stimuli are experienced less strongly than stimuli caused by external factors [53]. The strength with which a stimulus is experienced, therefore, can correlate with self-attribution of an action. *Temporal binding* refers to causally connected events being experienced as closer in time than unrelated ones. In the context of the sense of agency, this refers to the phenomenon where individuals perceive a closer temporal relationship between their actions and the resulting sensory outcomes than is objectively the case [36]. Measuring the experienced time interval between an action and its result can provide indicators of agency. If the user underestimates the two events systematically, one can assume temporal binding is experienced as well as a causal connection between the user's action and the result [22]. The two most common ways of measuring those experienced time intervals are the *Libet clock* and *interval estimation*.

The Libet clock shows a clock face without any numbers and a clock hand. The participants are asked to observe the rotating clock hand during the experiments and to report its position during the action taken and an observed outcome. This method requires the visual attention of the participant, which means the visual attention cannot be drawn to another visual cue during the experiment.

Using the interval estimation method, the participants are asked to report their estimate of the interval lengths (typically in milliseconds) [27]. This method is used in the study presented here, as it does not require the visual attention of the participant, which is needed for the images shown during the experiment.

As an implicit measure of the sense of agency, the literature often refers to the phenomenon of temporal binding as *intentional binding*. This is the term originally used by Haggard et al. [36], to highlight that the intentionality of the action seemed to be important for this phenomenon. *In the present study, we use temporal binding to estimate the pre-reflective sense of agency. We use interval estimation to prevent conflicts of visual attention.*

2.2 Generative Models

GenAI is based on generative models [39], which learn to capture the distribution of samples in a training dataset to produce a practically unlimited number of new samples that follow this distribution. This requires the model to mimic the potentially highly complex structures and correlations in the data. As an example, consider a large training corpus of photos of lions. A generative image model successfully trained on this data is able to synthesize new images that depict lions in a photo-realistic way. These images include lions in different poses, some of which intricately correlate with, e.g., the time of day (lions are more active at night). To capture these complex correlations, generative models tend to implicitly learn and represent useful, semantic knowledge about the data [9, 30], e.g., lions sleep during the day. This intriguing property makes them an appealing tool for interacting with content, since one can operate on the basis of higher-level concepts rather than use lower-level signal processing.

2.2.1 Generating Images. The concept of generative modeling is old [60], but only with the ongoing advancements in deep learning [34], it is now possible to create powerful models that are useful outside the research lab or rather simple use cases [30]. While a larger family of deep generative model types has been developed, a significant amount of current machine-learning research revolves around a small number of alternatives: *Diffusion models* [41, 74, 77, 78, 86, 88] mimic the data distribution using gradual de-noising; *Autoregressive Models* generate samples piece by piece, where each piece is conditioned on the already generated content [16, 33, 71, 75, 93]; and *Generative Adversarial Networks* (GANs) rely on a two-player game between a generator and a discriminator [35, 45–48, 79, 80]. In principle, generative models can operate on any data modality, as long as training datasets are large enough. However, recently, a lot of attention has been paid to text [16], images [74, 78], and videos [13, 42]. *In this work, we focus on image generation.*

2.2.2 Controlling the Generation. Synthesizing realistic, yet *random*, images is not particularly useful. Typically, users want to specify what the synthetic image should depict or, more generally, interact with the GenAI to produce custom content. Technically, there are multiple ways to achieve this goal. A popular approach is conditioning and guidance, i.e., users provide an additional control signal that steers the generation. Popular control modalities include class labels [14, 25], text [15, 28, 79], and scribbles [7, 63]. Unfortunately, such guidance signals are coarse-grained, and the user only has a somewhat fuzzy and indirect influence on the generation. Already existing (image) content can be used for conditioning as well [43, 103], but severely limits creative freedom and expressivity.

Another line of work focuses on operations in the latent space of the generative model [44], which oftentimes involves an inversion step [2, 3, 76, 87]. In this latent space, simple decompositions are surprisingly effective in finding interpretable directions to achieve image editing goals [38, 82]. It is also possible to learn more complex trajectories [4], and specialized solutions exist for narrow use cases, such as rotating faces [54, 92]. Also, for these approaches, editing and generation granularity and/or diversity are severely limited and do not allow full access to the broad scope of GenAI.

In contrast, geometric edits are more tangible and general. In particular, point-based manipulations allow direct control over which image location moves where [26, 68, 95]. DragGAN [68] introduced editing using an interleaved scheme of motion supervision and point tracking that allows geometric edits with unprecedented precision at interactive rates. Similar capabilities, yet without the interactive speed, have recently been explored for diffusion models as well [55, 66, 85]. *In this work, we focus our attention on DragGAN for its ability to provide users with precise, i.e., pixel-accurate control over the image generation process, while operating at interactive rates. This appealing combination, exposing prototypical idiosyncrasies of generative models, thus provides a meaningful testbed for our investigation of agency with GenAI.*

3 STUDY RATIONALE

This study aims to provide a foundational understanding of how users experience agency when a GenAI algorithm provides an output that has features not requested by the user. We are specifically focusing on GenAI image generation, and observing how different visual properties of the generated images effect agency.

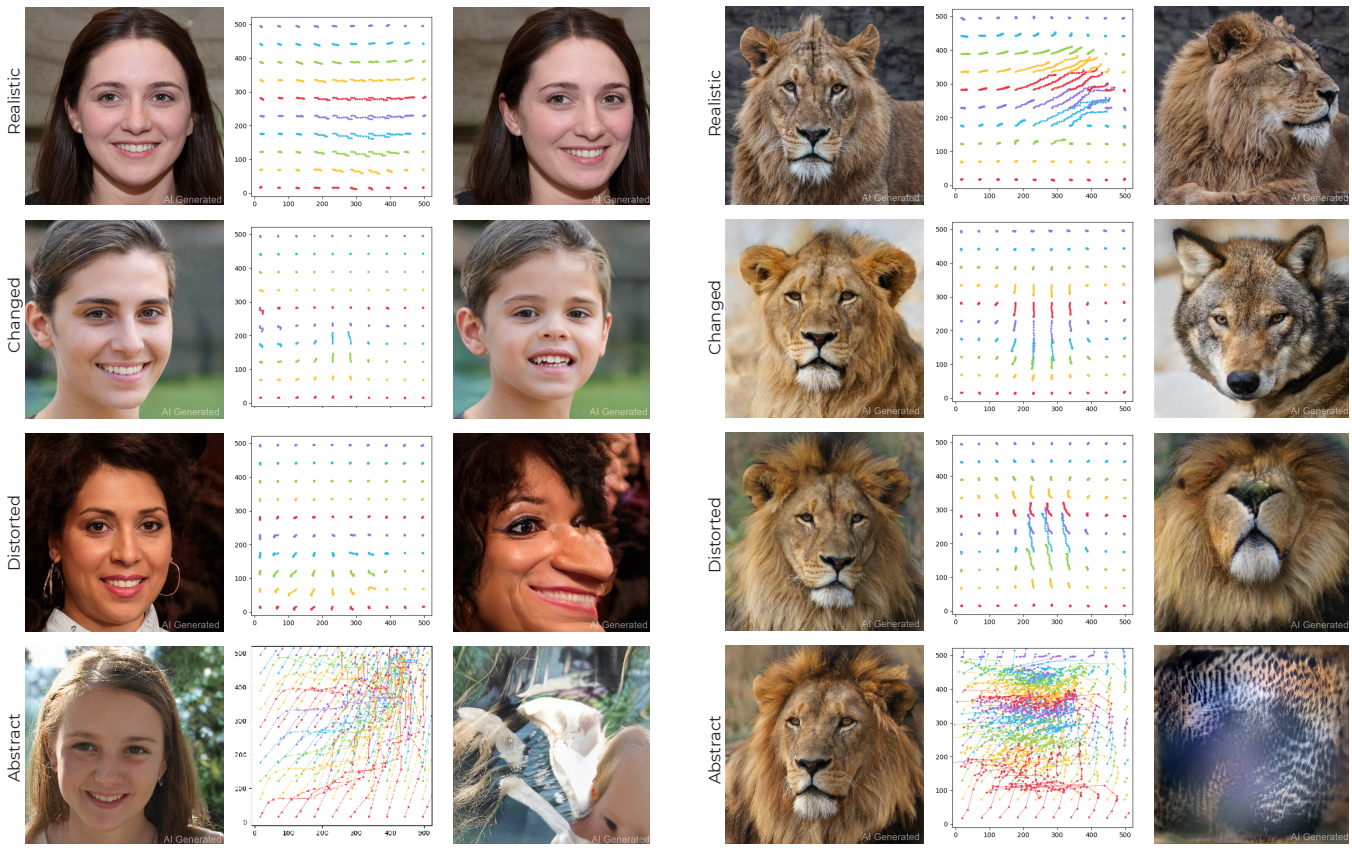

Figure 4: Four different pairs of starting and target images. From top to bottom: *realistic*, *changed*, *distorted*, and *abstract*. In between the starting image and target image, we show the visual flow computed using RAFT [91].

We consider this scenario relevant for understanding interactions where GenAI provides creative input in a shared process. Understanding how such interactions affect agency will provide a foundation for future GenAI specific interfaces.

3.1 Why DragGAN

We intend to measure both the pre-reflective sense of agency and user judgments of agency. Capturing the pre-reflective sense of agency is difficult; it cannot be directly measured, rather it must be inferred by other implicit measures. According to current literature, this requires specific conditions to be met, such as having simple and straight-forward tasks with one discrete point at which the user takes action and immediate feedback that can be manipulated in time [11, 21, 36, 64].

Furthermore, to better compare GenAI outputs that meet user expectations to outputs that differ from user expectations, we require an algorithm for which we can fine-tune to generate images with controlleable visual properties.

We have chosen DragGAN [68] as the approach that meets our criteria. DragGAN allows users to edit images synthesized by a GAN [35] using point manipulations. Provided with one or multiple start-target point pairs, the system optimizes a sequence of images where the image contents at the start point(s) move to their corresponding target point(s). Different from image warping, all this happens while images stay on the manifold of realistic images. For example, if a nose should move left, the face must follow, possibly revealing previously occluded parts of the face or the background and changing shadows on the face to account for the change in angle relative to the sun. These editing capabilities are achieved by leveraging the prior knowledge trained into a GAN. Any pre-trained GAN model can be used for this.

Depending on the pre-trained model, DragGAN also has other interesting properties that we use for our study. If a target position is correlated with another feature, dragging can lead to a change in that feature. For example, in the case of a model trained on faces of wild animals, if the starting image is a lion, and the training data contains many images of foxes that have their noses at the bottom of the image, dragging the lion's nose to the bottom can turn the lion into a fox. Furthermore, if the user drags a point to a space that the model has no training data for, the image will initially distort and finally completely warp into an abstract representation.

3.2 Research Questions & Image Qualities

Our primary research question asks *"How is the human experience of agency effected when GenAI provides an unexpected output?"* To

answer this question, we compare the experience of an image manipulation that produced a *realistic*, expected, target image to an image manipulation, where DragGAN *changes* salient features of the image, while preserving the interaction.

We ask a follow-up question "If there is a difference, is it because the new image is semantically meaningful, or is it simply because the image is different?" To answer this question, we introduce a new type of image, where rather than a meaningful transformation, the DragGAN *distorts* the image. Here, the final image still has some resemblance to the original image.

Finally, we ask "Is there a difference between pictures with clearly distinguishable semantic meaning and chaotic generation results?" To address this question, we add a final image class, where we push DragGAN to its limits, resulting in chaotic images without any discernible semantic meaning.

We use the following definitions for the four categories:

Realistic. The animal or human from the original image is still recognizable, and the change from the original image is a realistic movement (e.g., turning the head to the right). It also matches the position clicked on in the original image.

Changed. The animal or human in the resulting image looks realistic and matches the position the participant clicked on in the original image. However, it does not match the animal or human seen in the original image. For example, a lion turns into a fox, or an adult turns into a child.

Distorted. The lion's or the human's face becomes distorted when moved to the point the participant clicked on. The position still matches the position selected to some extent, but the image does not change as expected.

Abstract. The original image is not recognizable at all anymore, and the resulting image seems to show random shapes, colors, and textures. These tend to be based on the colors and textures in the original image. For example, in the human cases, the hair texture or clothing color can still be visible. In the lion cases, the fur texture and background colors tend to be recognizable. But, there are also cases, where none of this is recognizable anymore.

Examples can be found in Figure 4.

4 IMPLEMENTATION

In this section, we present the implementation details of our experiment, including the methodologies for image generation, the computation of image metrics, and the experimental setup.

4.1 Image Generation

For this study, we restrict ourselves to a single point used for dragging. We need to be able to control the latency between the participant's interaction with the system and the resulting image. With DragGAN, this is not possible in real time. The further the distance between the start and the end point, the longer it takes until the final image is created and shown. Different models and image resolutions also have an influence on the time needed until the final image is created. Therefore, we decided to pre-render selected images. For this, on each image, we selected a start point (usually around the nose of the animal or human) and used DragGAN to render the final image for possible end points. For most images, the result images of two adjacent pixels only show minor differences, so we limited the scope of this by only rendering the resulting image for every 8th pixel in each dimension. All images had a resolution of 512x512 pixels, which means for every starting image, 4096 possible resulting images were pre-rendered with DragGAN.

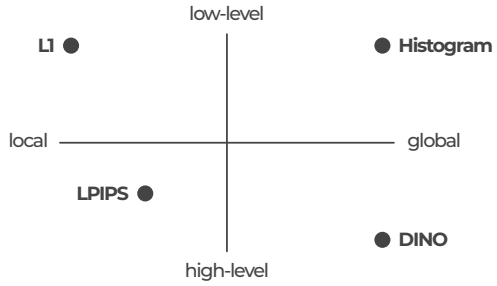

Figure 5: The chosen metrics describe image (dis)similarities based on low- and high-level features and consider local or global features.

Using the pre-rendered images has two advantages: First, users can freely and interactively drag the pre-selected image feature (for example, a nose), and second, we, the experimenters, can instrument delay times and other parameters.

Having pre-rendered resulting images also allowed us to run the study independently of the DragGAN software.

The images were selected from four different pre-trained StyleGAN2 models:

- stylegan2-afhqwild-512x512 - Lions with *changed* (step size = 0.001), *distorted* (step size = 0.001), and *abstract* (step size = 0.1) qualities.
- stylegan2_lions_512_pytorch - Lions with *realistic* (step size = 0.001), *distorted* (step size = 0.001), and *abstract* (step size = 0.06) qualities.
- stylegan2-ffhq-512x512 - Humans with *changed* (step size = 0.001), *distorted* (step size = 0.002), and *abstract* (step size = 0.1) qualities.
- stylegan2-ffhqu-1024x1024 - Humans with *realistic* (step size = 0.002), *distorted* (step size = 0.005), and *abstract* (step size = 0.1) qualities.

Appendix A shows Table 1 listing the properties of all 48 starting images that have been selected for pre-rendering, describing the image seed within the model and the starting point, from which the images have been rendered. For each species, two models have been chosen, as it was extremely difficult to find images with *changed* and *realistic* qualities within the same model. The starting images were selected by manually exploring the options within each model and picking starting images which resulted in target images featuring the desired visual properties.

4.2 Objective Metric Calculation

To place the magnitude estimation and temporal binding results in context, and to describe the differences between the image qualities in a quantitative way, we computed several metrics to evaluate

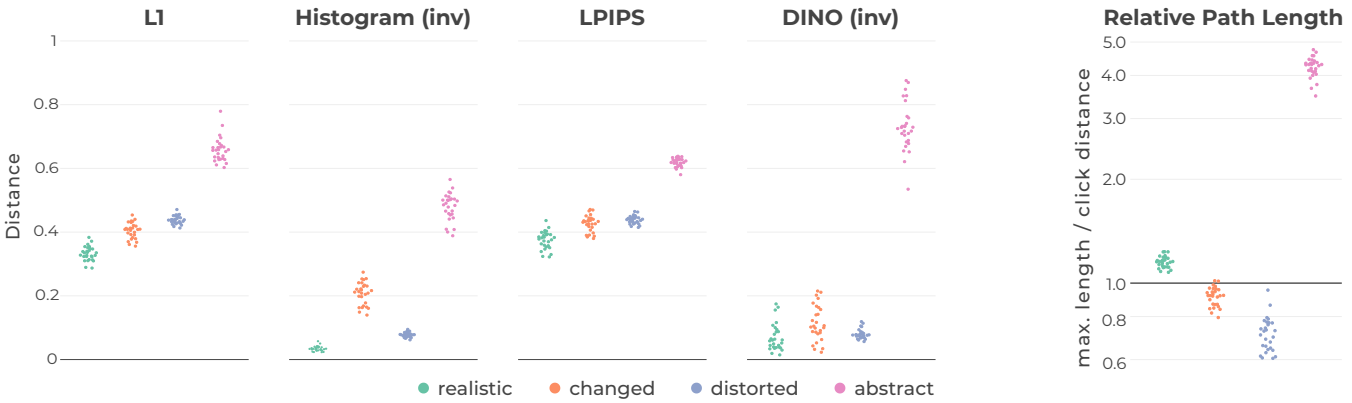

Figure 6: Distances calculated with four metrics (i.e., L1, histogram, LPIPS, and DINO) between the starting and the target images for each condition, and ratio of the longest trajectory (i.e., max. length) and the distance between user clicks (i.e., click distance).

differences between original and modified images: per-pixel difference (L1), correlation of RGB histograms (Hist.), a perceptually calibrated distance metric based on learned features (LPIPS [104]), and distance of features from a self-supervised vision transformer (DINO [91]).

The image metrics were selected to cover a broad spectrum of qualities. In particular, the metrics can be arranged on a 2D continuum that spans a space between local (L1, LPIPS.) vs. global (Hist., DINO) and low-level (L1, Hist.) vs. high-level (LPIPS, DINO) features, as shown in Figure 5. While local metrics consider differences in image content with higher spatial selectivity, global metrics aggregate information across the image before computing distances. Metrics based on low-level features operate close to the original signal (i.e., pixel values), while those based on high-level features first lift the signal to a more abstract representation, which typically allows to operate on a more abstract/semantic level.

Details of the calculations follow:

L1. We computed per-pixel absolute differences in RGB color space and average the result across all pixels. This provides a local and low-level estimate of how colors change.

Histogram Correlation. We computed a 3D histogram by assigning pixels to one of 8x8x8 regularly spaced RGB color bins. The distance between two histograms is determined by computing the Pearson correlation coefficient. This metric describes a shift in global color distribution. We inverted this measure to transform it into a difference metric.

LPIPS. This metric first feeds each image into a convolutional neural network (CNN) that has been pre-trained on large-scale data to perform image classification. The intermediate features (internal activations) of this network have been shown to be remarkably expressive: They have the emergent property that distances between images computed in this feature space strongly correlate with human perception. Based on this observation, the LPIPS [104] metric employs an additional calibration to match human perceptual judgment even better. By design, this metric operates on a range of features that correspond to different levels of abstraction: Features from earlier layers correspond to rather low-level information—Do edges in both images align?—while features from deeper layers tend to encode semantic concepts—Do we see the eye of a lion at the same location in both images?

DINO. This metric considers the cosine similarity between DINO [19] features. DINO uses a teacher–student model where both models turn an image into a 384-dimensional feature describing that image. During training, billions of images and many different crops of each image are shown to the models. While the student gets to see all the data, the teacher only sees a subset of crops. Additionally, the teacher is forced to learn more slowly than the student. The core objective in DINO is to make the student model's features similar to the ones of the teacher model for the same image, while ensuring diversity across different images. This construction creates a feature space where similar images have similar representations, i.e., a a rather detailed "understanding" of the visual world is obtained: Features of an image containing a pelican are vastly different from images containing a house, but only slightly different from images containing a tucan, yet still different enough to enable a clear distinction between pelicans and tucans. Measuring similarity of DINO features therefore gives a global, high-level estimate of differences in image content. As with histogram, we inverted this measure to transform it into a difference metric."

The raw data of each objective metric for each condition is provided in Figure 6. Scatterplots showing low-level, high-level, local, and global metrics are shown in Figure 7.

4.3 Optical Flow

To complement the above metrics, we also calculated optical flow. This investigates image differences in terms of *motion*: How does the original image need to be deformed to get the target image? Notice that, in contrast to the other metrics discussed above, this analysis is primarily focused on the image domain (position) rather than its co-domain (color). Humans have been shown to perform reasoning on the level of image deformations [83], and corresponding image

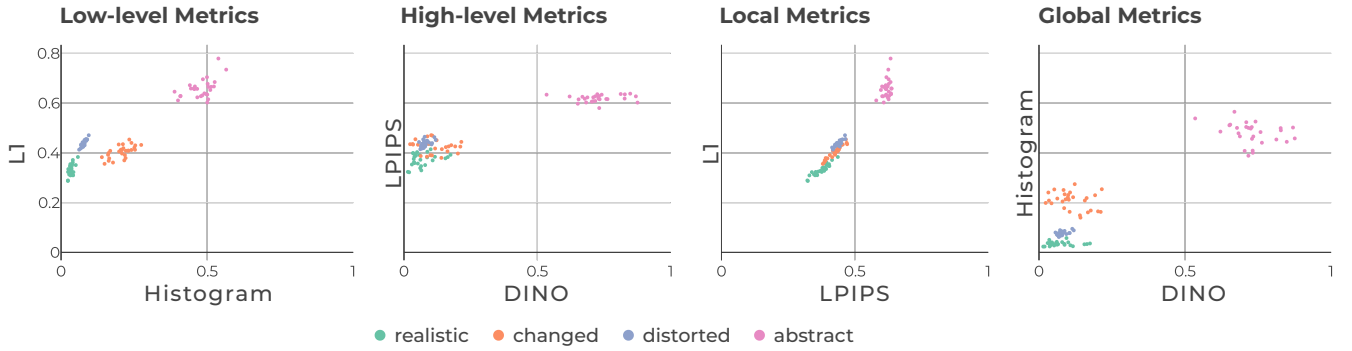

Figure 7: Comparison of distance measurements based on metrics categories. While all categories can clearly separate image pairs of the *abstract* condition, other conditions are not always linearly separable.

metrics have been explored [51, 52]. Analysis of motion is a natural choice for the geometric approach of DragGAN.

We use the state-of-the-art optical flow estimator RAFT [91] to compute per-pixel motion trajectories based on image sequences, from the original image to the user-specified target image. Specifically, we compute the optical flow between every two adjacent images and concatenate the flow vectors to yield full motion trajectories. To avoid clutter, our visualizations (Figure 4) show only a subset of trajectories. An index showing the ratio of click distance to trajectory length is shown in Figure 6 on the right.

4.4 Experiment Implementation

The study was implemented as a web application to make it accessible to the participants through a web browser on their PC or laptop. For the implementation, the Javascript library jsPsych [23] was used. This library helps in structuring the experiment and collecting the data. In this web application, the starting image was shown on an HTML Canvas element, with a half-transparent red dot marking the starting point and half-transparent green rectangles marking the clickable areas on the image. When an end point was selected, the resulting image for the closest pre-rendered end point was shown.

The participants were recruited and paid through Prolific[1]. Prolific linked to the web application hosting the experiment. Once participants completed the study, they were shown a code, which they entered on Prolific to receive payment. This additionally allowed the experimenters to access the participants' demographic data available on Prolific. This means that the web application itself collected no demographic data. Only the participants' informed consents and the study data were collected.

The web application was hosted on a server provided by the first author's home institution. The collected data was also stored on this server.

5 EXPERIMENT

Temporal binding and magnitude estimation were used to evaluate the influence of the qualities of the outcome image on the implicit feeling and the explicit judgment of agency. Temporal binding describes the phenomenon when people perceive an action and an outcome as closer to each other in time than they objectively are, which means that people perceive the time between the action and outcome as shorter. In this study, this measure is used as an implicit measure of the sense of agency by asking the participants to estimate the time interval between their action and the outcome. For this, the action and the outcome should be at discrete points in time to have an unambiguous interval length between them. Using interval estimation as a way to measure temporal binding is an established method that is widely used [6, 24, 102]. This method allows the participants to fully use their visual attention for the images presented throughout the study. Even though it is difficult for humans to precisely differentiate between intervals in milliseconds, we can see if systematic differences in the estimation errors are present throughout the different conditions [27].

Magnitude estimation is a method where the a participant assigns a numerical value to the perceived strength of a certain stimulus [59]. In this study, this measure is used as an explicit measure of the judgment of agency.

To control the image qualities as a condition, we highlighted pre-selected regions that would produce the desired visual properties with a green box (visible in Figure 8, leftmost image, top). Participants were instructed to only click on a spot on the image within these areas. For the lions, those areas summed to a total of around 40k sq. pixels, split into two areas per image. For the human images, it was more difficult to force certain outcome qualities, which resulted in one smaller area per image. This is a limitation due to the nature of the models these images were picked from.

5.0.1 Participants. Twenty-eight participants (12 male, 16 female) aged 20 to 49 (M = 32.36, SD = 7.51) took part in the online study. All participants were residents in the United Kingdom at the time of the study, were fluent in English, and had normal or corrected vision.

Three participants were removed prior to data analysis, as their data indicated non-compliance with instructions. We recruited an additional three participants. The demographics above are only of those participants whose data was used.

[1]prolific.co [accessed August 2023]

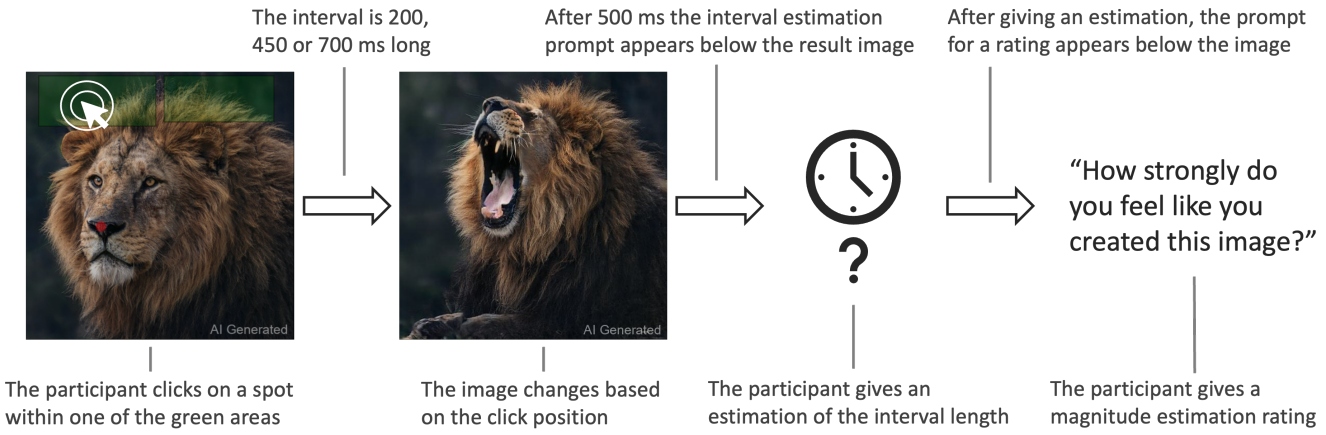

Figure 8: The procedure of the main part of the online study. The example image manipulation shows a starting image with a lion, the red starting point on its nose, and two green clickable areas at the top of the image. It also shows an outcome image based on the click position. The example is from the *realistic* condition. After the manipulation, the participants gave an estimation of the interval between their click and the appearance of the resulting image, to measure temporal binding, as an implicit measure for agency. They also rated how strongly they feel like they created this image explicitly.

5.0.2 *Procedure.*

On-boarding and training: The study was run online. It started by providing general study background and task information. After signing the consent forms, the participants began with a training session on estimating interval lengths in milliseconds. For this, a black circle outline was shown. After 50 to 850 milliseconds, the circle outline turned into a full black circle. The participants were asked to estimate the amount of time the outline was present before it changed to a full circle. Then, they received feedback telling them how long the interval really was. This procedure was done nine times, with 50, 100, 200, 350, 450, 550, 700, 800, and 850 millisecond intervals in random order. These interval lengths were chosen to cover the full range of options presented to the participants when asking for their estimation. The same task was repeated nine more times, without the feedback at the end, as a baseline for participants' interval estimates. In this situation, no temporal binding was expected, as there was no action taken by the participants. A second baseline measure was taken where temporal binding was expected. For this, again a circle outline was shown to the participants. This time, they had to click on the circle outline to change it to a full circle. The participants were asked to estimate the time interval between their click and the circle changing. This was repeated nine times, too. In both baseline conditions, each interval was 200, 450, or 700 milliseconds long, as those were also the interval lengths used in the main part of the study.

Main Task: After completing the training and the baseline measures, participants continued to the main part of the study: manipulating images. In each round, participants saw an AI-generated image, either a lion or a human. On each picture, we added a red circle to indicate the starting point of the manipulation and one (humans) or two (lions) green areas to indicate where on the image the participants could click to change the image. These areas were necessary to be able to control the visual properties of the generated image the participant saw in each round, as described in subsection 3.2.

After a time interval of 200, 450, or 700 milliseconds, the image changed based on the position of a participant's click and based on the visual properties selected for the current condition. After the manipulation, a question popped up below the resulting image, prompting each participant to give an estimation of the interval between their click and the resulting image appearing. The question presented possible response choices from 50 to 850 milliseconds in 50-millisecond steps. After selecting their estimation and clicking the "continue" button, they were prompted to rate how strongly they felt that they had created this image. This procedure can be seen in Figure 8. In total, there were 48 rounds of image manipulation - 24 with humans and 24 with lions. For each species, there were six images in each of the four image qualities category, which creates eight conditions with six images each. A visualization of these eight conditions is shown in Figure 4. The species and outcome qualities were balanced using a balanced Latin square.

6 RESULTS

In this section, we provide an overview of the generated images, results from measures of agency, and contextualize these results within the scope of our study.

6.1 Overview of Generated Images

The target regions used in the experiment were based on a subjective evaluation of the resulting images they produced. To highlight that these regions did indeed lead to clearly distinct clusters, we show the resulting distance measures per quality in Figure 6.

We show scatterplots of low-level, high-level, local, and global metrics plotted against one another (Figure 7). The scatterplots show clear linear separability for low-level and global metrics. It is

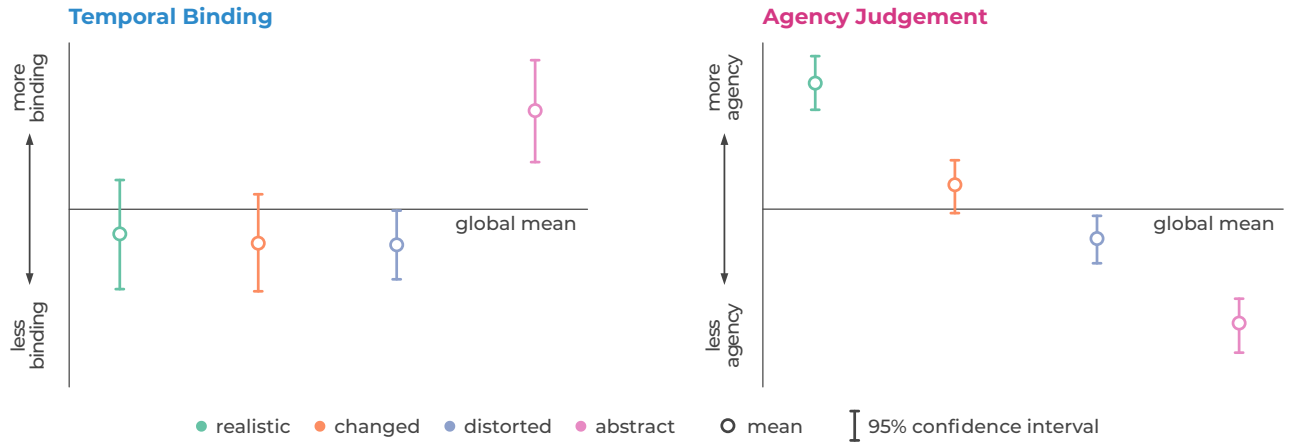

Figure 9: Left: Standardized binding scores for the four different conditions of image qualities. The circles show the mean, and the lines show a 95% confidence interval. There is a significant difference between *abstract* and *changed* as well as between *abstract* and *distorted*. Right: Standardized magnitude estimations for the four conditions of image qualities. The circles show the mean, and the lines show a 95% confidence interval. The differences are significant between all four conditions.

of note that the high-level metrics are not able to clearly distinguish between *realistic*, *changed*, and *distorted* images.

6.2 Measures of Agency

6.2.1 Temporal Binding. The binding scores to evaluate temporal binding were calculated by subtracting the interval estimate from the respective actual interval length. Positive binding scores thus indicate an underestimation of the interval and with this, temporal binding for this interval.

To identify significant differences between the conditions, the binding scores were standardized for each participant by first removing the participant's average response from each estimate and then dividing each estimate by the standard deviation. Resulting data is in standardized units. An estimate of one indicates that this estimate is one standard deviation above the average estimate for that user. This highlights differences between conditions, while removing individual differences between participants. We then performed a repeated measures ANOVA with qualities and species as independent variables and standardized binding scores as a dependent variable. Figure 9 shows the standardized binding scores for each condition with a 95% confidence interval.

There was a significant effect of qualities, $F_{(2.386,64.422)} = 5.312$, $p = .005$, $\eta_p^2 = .164$. There was no significant effect for species, $F_{(1,27)} = .394$, $p = .536$, $\eta_p^2 = .014$, and the interaction of species and quality $F_{(3,81)} = .619$, $p = .605$, $\eta_p^2 = .022$. Figure 9 (left) shows that the *abstract* condition received a higher average score than all other conditions. A Bonferroni corrected post-hoc comparison indicated that the difference between *abstract* and *distorted* was significant ($p < .001$), as was the difference with *changed* ($p = .024$). Though visual inspection appears to suggest a difference between *abstract* and *realistic* this is not possible to confirm statistically, due to correcting for multiple comparisons.

6.2.2 Magnitude estimation. A further repeated measures ANOVA was done with qualities, and species as independent variables and magnitude estimation as a dependent variable. As before, the estimation scores were standardized for each participant. We found a significant effect of species, $F_{(1,27)} = 4.963$, $p = .034$, $\eta_p^2 = .155$, and quality, $F_{(1.935,52.244)} = 42.987$, $p < .001$, $\eta_p^2 = .614$, but no significant effect of the interaction thereof, $F_{(3,81)} = 1.389$, $p < .252$, $\eta_p^2 = .049$.

Figure 9 (right) shows the average and 95% confidence intervals of these standardized estimates. The figure shows *realistic* received the highest rating, followed by *changed*, *distorted*, and, finally, *abstract*. A Bonferroni corrected Post-hoc test indicated that, except for *changed* and *distorted*, all difference are statistically significant at the p < .001 level.

6.3 Contextualization of results

To provide a qualitative intuition of what might be the underlying properties of the generated images that caused the temporal binding and agency judgment results, we provide visualizations of these results correlated with the image difference measures we calculated (Figure 10). To understand how strongly the movement varied between visual properties of images, we provide the path length as a function of the click distance. If the resulting value is one, then the movement matched the distance requested by the user higher or lower numbers indicate longer or shorter paths than one would expect based on click-distance alone (see Figure 6, right).

It should be noted that this metric becomes more relevant in future work which will investigate continuous input, and is presented here for completeness and future reference.

7 DISCUSSION

The experimental results (Figure 9) show clearly distinct patterns between temporal binding results and agency judgments. The temporal binding task (Figure 9, left) produced two distinct clusters – *realistic*, *changed*, and *distorted* images with less binding and *abstract* images with more binding. The agency judgments produced

in the magnitude estimation task (Figure 9, right) show highest agency with *realistic* images and then increasingly reduced agency for *changed*, *distorted*, and, finally, *abstract* images.

That the results are different is not surprising as both experiments estimate different aspects of agency, as discussed in our related work; however, it is interesting to reflect on what might have caused the specific patterns we observed.

7.0.1 Initial Reflection.

Explaining Temporal Binding. Our experimental method, interval estimation, is not ideal for determining *if* there was temporal binding[2]. Instead, the results (Figure 9 , left) should be interpreted as differences in experienced binding across conditions. Literature suggests participants likely experienced temporal binding in all conditions [64, 100].

Why is binding significantly greater in the *abstract* condition and similar in the *realistic*, *changed*, and *distorted* conditions? Since the participants' actions did not change, we must consider the outcomes. Studies in cognitive psychology have shown that intention and agency are not the only potential causes of temporal binding. Factors such as causal relations, attention, and arousal can affect time perception and influence binding [100]. We can dismiss causal relations, as all conditions involved them; however, attention and arousal do offer plausible explanations for the difference in the *abstract* condition.

Explaining Agency Judgments. Magnitude estimation results suggest a continuum of how closely output images match user expectations, from *realistic* to *abstract*. The closer the image matches user expectations, the higher the rating appears to be, aligning with Wegner's theory of agency, particularly consistency [98]. The *realistic* condition produced outcomes most consistent with participants' expectations. Consistency decreased in other conditions: In *changed*, the face rotated and changed age or species; in *distorted*, the image became warped but recognizable; in *abstract*, the image became unrecognizable.

One interpretation of this data is that users reflected less on their internal state while perceiving the image, but instead on the image itself. Users might have estimated how much the target image differed from the original image and provided their judgment accordingly.

The calculated objective metrics provide further data for understanding the results of the subjective measures of pre-reflective agency and agency judgments, which we will explore in the next section.

7.1 Hypothesizing with distance metrics

We provide both subjective measures of human experience—*temporal binding* for pre-reflective agency and *magnitude estimation* for agency judgments —as well as calculated calculated objective image metrics—L1, Histogram, LPIPS, and DINO. This allows us to formulate tentative hypotheses and estimate if they are supported by the data. To provide a visual overview of how the subjective measures and objective metrices relate, we provide plot all subjective and objective measures in Figure 10.

Next, we will discuss four hypotheses we found useful in understanding the results. To avoid getting lost in details, we will discuss the hypotheses primarily using Figures 6 and 7.

H1 *Pre-Reflective Agency is based on raw image differences.* One method to characterize the changes in the image according to user input is to quantify the modifications in the pixel values. On a global level, we can measure changes by looking at the histogram of the image, and, locally, we can measure changes based on the difference in color of individual pixels. The temporal binding measure was unable to capture differences between *realistic*, *changed*, and *distorted* conditions, but it highlights that the *abstract* condition was clearly different (Figure 9). To claim that *H1* is supported by the data, we might look for similar patterns in the calculations of low-level image metrics such as L1 and Histogram (Figure 6). While the low-level calculated measures show that the *abstract* condition is different from the other conditions, these measures also distinguish between the *realistic*, *changed*, and *distorted* conditions (see scatterplot Figure 7, Low-level Metrics), which is not reflected in the subjective measures. We argue that *H1* is *not supported* by the observed data.

H2 *Pre-Reflective Agency is based on semantic image differences.* User input also has another effect: the content of the images changes. A face might look in another direction, transform, or vanish altogether. This changes the semantic content of the image, which is captured by the high-level metrics LPIPS and DINO (Figure 6). We can argue that the data supports *H2* if we find a similar clustering in the LPIPS and DINO calculations, which, indeed, we do (Figure 7, High-level Metrics). We argue that *H2* is *supported* by the observed data. However, looking towards the similarity of the *realistic* and *changed* condition, the difference needs to be sufficiently large to produce a measurable effect.

H3 *Agency Judgment is based on local image differences.* Unlike the two clusters we found in the temporal binding data, the magnitude estimation data appears to linearly relate to the different image properties, clearly capturing differences between all four conditions (Figure 9, right). As the literature suggests that Agency Judgment is based on higher-level cues, such as intentions, thoughts, and social context, we did not expect to find correlations to these results in the calculated measures. However, while not expected, we found a pattern similar to the magnitude estimation results when looking at the local measures L1 and LPIPS (Figure 6 and the scatterplot in Figure 7, Local Metrics). We therefore argue that *H3* is *supported* by the observed data.

H4 *Agency Judgment is based on global measures.* As a counterpoint to *H3*, we might hypothesize that agency judgment is also based on global measures. We argue that this is the case if we find a similar linear relation between agency judgments and calculated global measures Histogram and DINO (Figure 6); however, this is not found in the data. In fact, as can be seen in the scatterplots in Figure 7 (Global Metrics) and in Figure 10 (Magnitude Estimation vs. Histogram and DINO), this is not the case; instead, we found a non-monotonous relation. We therefore argue that *H4* is *not supported* by the observed data.

[2]This would require a Libet-clock-based task. We chose interval estimation to avoid presenting participants with competing visual tasks.

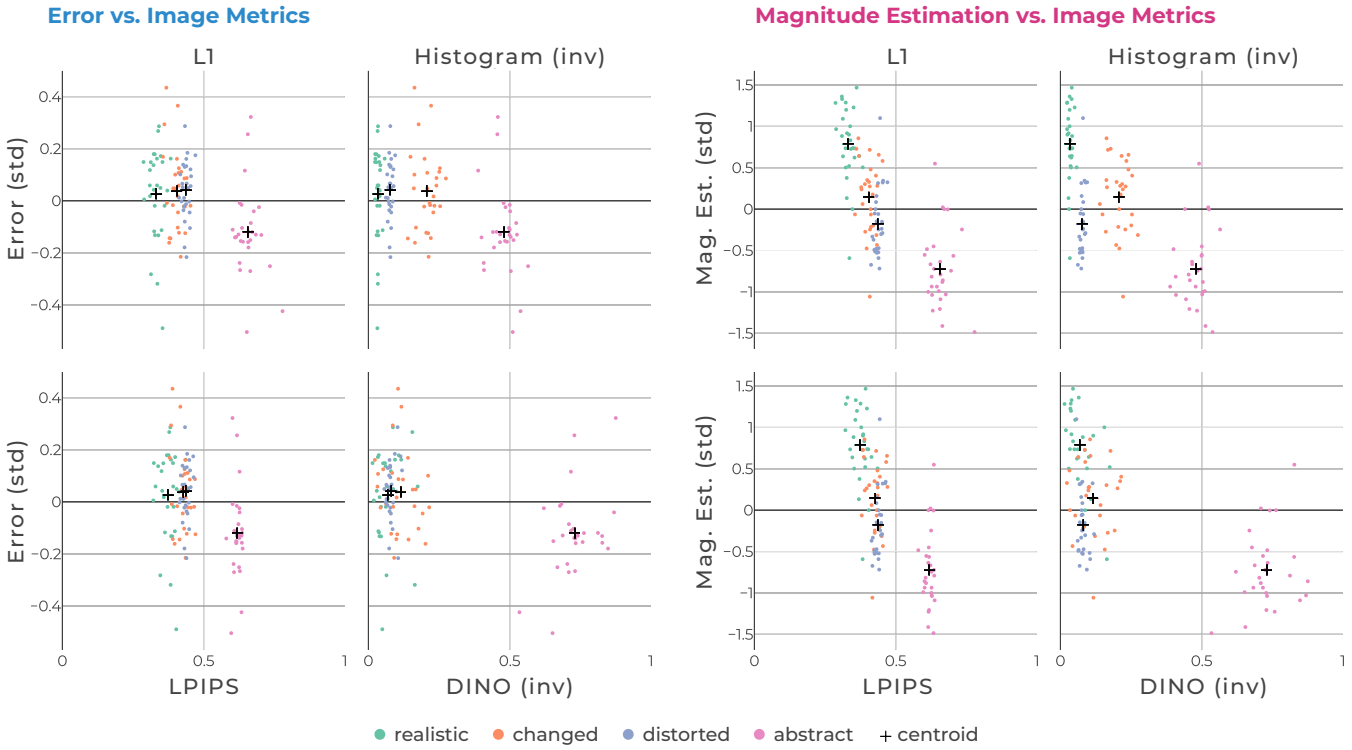

Figure 10: Scatterplots of objective measures of image differences plotted vs. temporal binding results and magnitude estimation results. The x-axes indicate the normalized differences of image pairs, i.e., higher values refer to higher dissimilarity.

In Summary (and a word of caution...) The hypotheses we discuss are merely a subset of potential questions we could ask, and surely do not adequately reflect the full complexity of our observations. However, they form useful take-home messages that might be used as a basis for future studies: The pre-reflective sense of agency correlates with semantic image differences, while agency judgments correlate with the magnitude of local differences between images. Having said this, correlation is not causation. We provide the distance metrics as a tool to hypothesize and theorize. We believe that this may provide a useful starting point for future experiments that can investigate these ideas in a causal manner. What we provide here is purely exploratory and should not be considered more than data-driven speculation.

7.1.1 But what about trajectories? The trajectories shown in Figure 4 (center) are a fundamentally different way of looking at the differences in visual properties. Here, we look at how features of the image move in space, rather than analyze the image in its color domain. This has been shown to closely match how humans think about images [83]. We found results that look seductively similar to the temporal binding results – short organized trajectories for *realistic*, *changed*, and *distorted* images and long chaotic paths for *abstract* images.

However, the *realistic* conditions tend to also exhibit longer paths than the *changed* and *distorted* condition, as DragGAN could fluidly move between starting and target images, even if we correct for actual distance between starting position and user click (Figure 6, right). For *changed* and *distorted* images, the paths we found tended to be shorter. This difference in flows was not observed in the experimental data.

Finally, we acknowledge that optical flow estimators are designed to estimate the (non-rigid) projected motion of real-world actions as seen in video sequences. In our RAFT-based framework, the analysis is conducted on pairs of adjacent frames. We observe that for the *realistic*, *changed*, and *distorted* conditions, changes in successive DragGAN-generated images can be consistently explained by optical flow. However, the chaotic trajectories in the *abstract* condition are often so incoherent that the optical flow estimator failed to identify correspondences between the images.

We believe that such spatial analysis of image differences will become more important in follow-up experiments that include continuous input tasks.

7.2 Reflective Judgement & Cognitive Dissonance Resolution

Driven by hunger, a fox tried to reach some grapes hanging high on the vine but was unable to, although he leaped with all his strength. As he went away, the fox remarked "Oh, you aren't even ripe yet! I don't need any sour grapes."

– Fable Attributed to Gaius Julius Phaedrus

While it is not surprising based on previous studies that have found similar effects, it is quite striking that the magnitude estimation results are in such stark contradiction to the temporal binding results. They are, however, also measuring very different things. While the temporal binding result is indicative of cognitive processes that happen while the interaction is occurring, the magnitude estimation occurs upon reflection on the outcome and on one's role in creating it. While the personal stories we tell ourselves do not feature in the temporal binding measures, they are an essential feature of the reflective evaluation that occurred in the magnitude estimation. Here, our need to create cohesive self-narratives is a driving force that shapes how we perceive the interaction. Much like the Fox who states that the grapes must be sour when he realizes that he cannot reach them, the users also adapt their responses to match their mental self model.

This way of adapting personal narratives is an essential part of human cognition. In fact, it is so ingrained, that studies have shown even people who are unable to make new short term memories and, therefore, are unable to remember what caused them to create such post-hoc biases, display the tendency to adjust their beliefs to fit such narratives [56].

The narrative view of agency [98] may prove particularly relevant in understanding human-AI interaction and GenAI in particular. In simple interactions with non-intelligent objects, exclusivity is framed around the question 'Did I do that?' The choice is binary. Either I did it, or I did not. The question becomes more complex when we interact with AI. Now it is not enough to ask 'Did I do that?' We can also ask 'Did the AI do that?' Inconsistency and the lack of exclusivity have the potential to diminish people's instinctive feeling of agency. In such situations, an explanation and a reflective judgment regarding agency is required. People consider their role and the role of the AI. Their judgment is based on the story they can plausibly tell, rather than a pre-reflective innate sense of agency.

7.3 Implications for Designing Interactions with GenAI

Currently, interfaces with GenAI usually deploy a *genie metaphor*. The GenAI is portrayed as an agent that we communicate with as an *other*, in strong contrast to the tool-metaphors otherwise prevalent in interacting with computers. With a few notable exceptions [58, 61], the details of how this is implemented are still often reminiscent of early command-line interfaces. This shows that there is still room for much innovation in the development of the interfaces we have for interacting with GenAI algorithms.

While HCI is continuously evolving, new technologies often spawn entire new subfields of research; for instance, "the rise of robots" led to Human Robot Interaction (HRI), a field distinct from HCI due to the physical nature of robots and their ability to interact with the physical world [99]. Similarly, we argue that the current developments in GenAI will require its own context-specific approach to interface design. Not only do we need new creative approaches towards using GenAI algorithms that go beyond command-line paradigms, these new interfaces will also need to deal with themes such as attribution and control.

Our study, once again, highlights the difference between the pre-reflective sense of agency and the judgment of agency. Because the sense of agency is purely subconscious and the judgment of agency is what we consciously have access to, a large body of HCI work [49, 50, 90] focuses on judgments. However, we argue that the judgments are based on the sense, modulated by one's context and personal narrative. As this personal narrative is typically outside the scope of what can be influenced by a simple interface, explicitly designing for the pre-reflective sense is often the most practical tool available for interface designers.

Here, our study highlights an opportunity for designers. While there appears to be an inherent contradiction between providing control to the user and allowing GenAI to provide creative and non-expected input, we did not find such a conflict in our data. We did not find any difference in temporal binding between the *realistic* and the *changed* conditions. This suggests that it is possible to maintain users' sense of agency while enabling GenAI to extend beyond users' specific requests.

7.4 Limitations and Future Work

The presented study focuses on GenAI for *image generation*, specifically a discrete direct image manipulation task. To understand the extent to which the results are specific to the chosen task and modality or generally true for interaction with GenAI, it needs to be contextualized in studies looking at other media (such as text generation or audio generation) and studies looking at other input types (such as continuous input or non-direct manipulation tasks). A key challenge here is identifying agency measures and experimental methods that can be used across such different contexts.

The present study already highlights some of the complexity in conducting such experiments. We used interval estimation as a way to measure temporal binding as an indicator for pre-reflective agency, as it fits the design of the experiment best. However, generally speaking, it is not an optimal measure. We also believe that the interface we used – radio buttons for selecting from a fixed list of interval times – introduced bias towards the middle (see also the interesting discussion by Matejka et al [62]). We suggest instead using a more classical magnitude estimation task, as we did for Agency Judgements to asess these intervals.

No matter what method used to collect estimates, the traditional method of using a Libet clock would still be preferred. It would allow the collection of details on when the binding occurs - for the action or outcome, or both. However, it would need the participants' full visual attention, which would have caused a problem with the images shown during this experiment. Alternatives to the Libet clock have been presented [20], which would also allow the distinction between action and outcome binding but do not use a participant's visual attention. A sensory attenuation approach is also promising, but identifying a useful implementation in this context is challenging.

The current study also focuses on a stripped down minimal application. To better understand the real-world relevance of different levels of agency, more complex scenarios, possibly where the user has an actual stake in the outcome, are required.

7.4.1 Follow-up work. We identify three main avenues for expanding upon this work. The first involves extending the current experimental approach to other media, such as text or audio. The second

avenue is investigating how different interface design choices influence agency; for instance, comparing non-direct with direct manipulation or discrete input with continuous input. These questions can likely be explored using quantitative methods, as demonstrated in the present paper, through methods suggested by Cornelio et al. [20], or using a sensory attenuation approach [53]. The third avenue is providing users with more complex tasks in which they have a personal stake. Qualitative methods, such as explicitation approaches [72], may be particularly useful in these scenarios.

We found the dual approach of calculating objective metrics and comparing them with subjective measures, as explored in our work, promising. Building on our exploratory analysis, researchers can formulate clear hypotheses for null-hypothesis testing or other forms of significance estimation. Such work will help build a better understanding of how to interpret the results of such experiments. This is important, as a crucial area for further development is not only understanding how GenAI design and use affect user agency, but also improving and developing robust tools and methods for assessing agency in this context.

8 CONCLUSION

This paper explored user experiences and perceived agency during GenAI content creation. Agency is closely linked to responsibility and ownership, which can blur during AI interactions. The paper examined how different qualities of AI-generated images influenced users' perceived agency. It showed that the participants felt a stronger sense of implicit agency when the AI produced *abstract* images, while they were more likely to, upon reflection, take credit for *realistic* images.

Most importantly, we found that small modifications created by the *changed* condition did not influence our measure of pre-reflective agency. This shows that, in principle, it is possible for GenAI to introduce novel elements without disrupting the subjective experience of agency.

When compared with computational measures of image differences, we found that temporal binding results seem to correlate with large semantic differences, whereas judgments of agency was better explained by local differences between images.

With this paper, we contribute towards considering agency as an active design dimension when designing GenAI interfaces. We provide an initial experiment and exploratory analysis, which we hope will act as a point of departure for future experimental work.

ACKNOWLEDGMENTS

We would like to thank Courtney N. Reed for her involvement in this project. Her early input shaped the direction of this work.

This work was conducted with financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224.

Support was also provided by Science Foundation Ireland through the Insight Centre for Data Analytics (12/RC/2289_P2).

REFERENCES

[1] n.d.. How Does Artificial Intelligence Transform Society? — futurelearn.com. https://www.futurelearn.com/info/courses/philosophy-of-technology/0/steps/256162. [Accessed 03-04-2024].

[2] Rameen Abdal, Yipeng Qin, and Peter Wonka. 2019. Image2StyleGAN: How to embed images into the stylegan latent space?. In *Proceedings of the IEEE/CVF international conference on computer vision*. 4432–4441.

[3] Rameen Abdal, Yipeng Qin, and Peter Wonka. 2020. Image2StyleGAN++: How to edit the embedded images?. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8296–8305.

[4] Rameen Abdal, Peihao Zhu, Niloy J Mitra, and Peter Wonka. 2021. Styleflow: Attribute-conditioned exploration of stylegan-generated images using conditional continuous normalizing flows. *ACM Transactions on Graphics (ToG)* 40, 3 (2021), 1–21.

[5] Marc Andersen, Kristoffer L Nielbo, Uffe Schjoedt, Thies Pfeiffer, Andreas Roepstorff, and Jesper Sørensen. 2019. Predictive minds in Ouija board sessions. *Phenomenology and the Cognitive Sciences* 18 (2019), 577–588.

[6] Zeynep Barlas, William E. Hockley, and Sukhvinder S. Obhi. 2018. Effects of free choice and outcome valence on the sense of agency: evidence from measures of intentional binding and feelings of control. *Experimental Brain Research* 236 (1 2018), 129–139. Issue 1. https://doi.org/10.1007/S00221-017-5112-3/FIGURES/3

[7] David Bau, Hendrik Strobelt, William Peebles, Jonas Wulff, Bolei Zhou, Jun-Yan Zhu, and Antonio Torralba. 2019. Semantic Photo Manipulation with a Generative Image Prior. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)* 38, 4 (2019).

[8] Florian Markus Bednarski, Kristina Musholt, and Charlotte Grosse Wiesmann. 2022. Do infants have agency? – The importance of control for the study of early agency. *Developmental Review* 64 (June 2022), 101022. https://doi.org/10.1016/j.dr.2022.101022

[9] Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35, 8 (2013), 1798–1828.

[10] Bruno Berberian, Patrick Le Blaye, Christian Schulte, Nawfel Kinani, and Pern Ren Sim. 2013. Data transmission latency and sense of control. In *Engineering Psychology and Cognitive Ergonomics. Understanding Human Cognition: 10th International Conference, EPCE 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part I 10*. Springer, 3–12.

[11] Joanna Bergstrom-Lehtovirta, David Coyle, Jarrod Knibbe, and Kasper Hornbæk. 2018. I really did that: Sense of agency with touchpad, keyboard, and on-skin interaction. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018). https://doi.org/10.1145/3173574.3173952 cognitive band; no co-actor(s).

[12] Joanna Bergström, Jarrod Knibbe, Henning Pohl, and Kasper Hornbæk. 2022. Sense of Agency and User Experience: Is There a Link? *ACM Transactions on Computer-Human Interaction (TOCHI)* 29 (3 2022), 1–22. Issue 4. https://doi.org/10.1145/3490493

[13] Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, et al. 2023. Stable video diffusion: Scaling latent video diffusion models to large datasets. *arXiv preprint arXiv:2311.15127* (2023).

[14] Andrew Brock, Jeff Donahue, and Karen Simonyan. 2018. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096* (2018).

[15] Tim Brooks, Aleksander Holynski, and Alexei A Efros. 2023. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18392–18402.

[16] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[17] Kalina Bryant. 2023. How AI Is Impacting Society And Shaping The Future — forbes.com. https://www.forbes.com/sites/kalinabryant/2023/12/13/how-ai-is-impacting-society-and-shaping-the-future/. [Accessed 03-04-2024].

[18] Daniel Buschek. 2024. Collage is the New Writing: Exploring the Fragmentation of Text and User Interfaces in AI Tools. *Designing Interactive Systems Conference* (7 2024), 2719–2737. https://doi.org/10.1145/3643834.3660681

[19] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. 2021. Emerging properties in self-supervised vision transformers. In *ICCV*. 9650–9660.

[20] Patricia I. Cornelio Martinez, Emanuela Maggioni, Kasper Hornbæk, Marianna Obrist, and Sriram Subramanian. 2018. Beyond the Libet Clock: Modality Variants for Agency Measurements. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174115

[21] David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. 2012. I did that! (2012), 2025. https://doi.org/10.1145/2207676.2208350 cognitive band; personal agencyExperiments on sense of agency1. Change of input modality, press button on keyboard and on own arm2. Aid of computer. Click on a circle, computer aided on different levels.

[22] David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. 2012. I Did That! Measuring Users' Experience of Agency in Their Own

Actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2025–2034. https://doi.org/10.1145/2207676.2208350

[23] Joshua R De Leeuw. 2015. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods* 47, 1 (2015), 1–12. https://doi.org/10.3758/s13428-014-0458-y

[24] John A. Dewey. 2024. Feelings of responsibility and temporal binding: A comparison of two measures of the sense of agency. *Consciousness and Cognition* 117 (1 2024), 103606. https://doi.org/10.1016/J.CONCOG.2023.103606

[25] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems* 34 (2021), 8780–8794.

[26] Yuki Endo. 2022. User-Controllable Latent Transformer for StyleGAN Image Layout Editing. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 395–406.

[27] Kai Engbert, Andreas Wohlschläger, Richard Thomas, and Patrick Haggard. 2007. Agency, Subjective Time, and Other Minds. *Journal of Experimental Psychology: Human Perception and Performance* 33 (2007), 1261–1268. Issue 6. https://doi.org/10.1037/0096-1523.33.6.1261

[28] Dave Epstein, Allan Jabri, Ben Poole, Alexei Efros, and Aleksander Holynski. 2024. Diffusion self-guidance for controllable image generation. *Advances in Neural Information Processing Systems* 36 (2024).

[29] Ziv Epstein, Aaron Hertzmann, Memo Akten, Hany Farid, Jessica Fjeld, Morgan R. Frank, Matthew Groh, Laura Herman, Neil Leach, Robert Mahari, Alex "Sandy" Pentland, Olga Russakovsky, Hope Schroeder, and Amy Smith. 2023. Art and the science of generative AI. *Science* 380, 6650 (June 2023), 1110–1111. https://doi.org/10.1126/science.adh4451

[30] David Foster. 2022. *Generative deep learning.* " O'Reilly Media, Inc.".

[31] Christopher D. Frith, Sarah-Jayne Blakemore, and Daniel M. Wolpert. 2000. Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 355, 1404 (2000), 1771–1788. https://doi.org/10.1098/rstb.2000.0734 arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2000.0734

[32] Chris D Frith, Sarah-Jayne Blakemore, and Daniel M Wolpert. 2000. Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews* 31, 2 (2000), 357–363. https://doi.org/10.1016/S0165-0173(99)00052-1

[33] Oran Gafni, Adam Polyak, Oron Ashual, Shelly Sheynin, Devi Parikh, and Yaniv Taigman. 2022. Make-a-scene: Scene-based text-to-image generation with human priors. In *European Conference on Computer Vision*. Springer, 89–106.

[34] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

[35] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *NeurIPS*. 2672–2680.

[36] Patrick Haggard, Sam Clark, and Jeri Kalogeras. 2002. Voluntary action and conscious awareness. *Nature Neuroscience* 5 (2002), 382–385. Issue 4. https://doi.org/10.1038/nn827 Intentional binding study/definition.

[37] Patrick Haggard and Manos Tsakiris. 2009. The Experience of Agency: Feelings, Judgments, and Responsibility. *https://doi.org/10.1111/j.1467-8721.2009.01644.x* 18 (8 2009), 242–246. Issue 4. https://doi.org/10.1111/J.1467-8721.2009.01644.X

[38] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. 2020. Ganspace: Discovering interpretable gan controls. *Advances in neural information processing systems* 33 (2020), 9841–9850.

[39] GM Harshvardhan, Mahendra Kumar Gourisaria, Manjusha Pandey, and Siddharth Swarup Rautaray. 2020. A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review* 38 (2020), 100285.

[40] James M Henslin. 1967. Craps and magic. *American journal of Sociology* 73, 3 (1967), 316–330.

[41] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* 33 (2020), 6840–6851.

[42] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. 2022. Video diffusion models. *Advances in Neural Information Processing Systems* 35 (2022), 8633–8646.

[43] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.

[44] Ali Jahanian, Lucy Chai, and Phillip Isola. 2019. On the" steerability" of generative adversarial networks. *arXiv preprint arXiv:1907.07171* (2019).

[45] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. 2020. Training generative adversarial networks with limited data. *NeurIPS* 33 (2020), 12104–12114.

[46] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2021. Alias-free generative adversarial networks. *NeurIPS* 34 (2021), 852–863.

[47] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *CVPR*. 4401–4410.

[48] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and improving the image quality of stylegan. In *CVPR*. 8110–8119.

[49] Shunichi Kasahara, Jun Nishida, and Pedro Lopes. 2019. Preemptive action: Accelerating human reaction using electrical muscle stimulation without compromising agency. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.

[50] Shunichi Kasahara, Kazuma Takada, Jun Nishida, Kazuhisa Shibata, Shinsuke Shimojo, and Pedro Lopes. 2021. Preserving agency during electrical muscle stimulation training speeds up reaction time directly after removing EMS. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–9.

[51] Petr Kellnhofer, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. 2015. A transformation-aware perceptual image metric. In *Human Vision and Electronic Imaging XX*, Vol. 9394. SPIE, 49–62.

[52] Markus Kettunen, Erik Härkönen, and Jaakko Lehtinen. 2019. E-lpips: robust perceptual image similarity via random transformation ensembles. *arXiv preprint arXiv:1906.03973* (2019).

[53] Fabian Kiepe, Nils Kraus, and Guido Hesselmann. 2021. Sensory Attenuation in the Auditory Modality as a Window Into Predictive Processing. *Frontiers in Human Neuroscience* 15 (Nov. 2021). https://doi.org/10.3389/fnhum.2021.704668

[54] Thomas Leimkühler and George Drettakis. 2021. FreeStyleGAN: Free-view Editable Portrait Rendering with the Camera Manifold. 40, 6 (2021). https://doi.org/10.1145/3478513.3480538

[55] Ruining Li, Chuanxia Zheng, Christian Rupprecht, and Andrea Vedaldi. 2024. DragAPart: Learning a Part-Level Motion Prior for Articulated Objects. *arXiv preprint arXiv:2403.15382* (2024).

[56] Matthew D. Lieberman, Kevin N. Ochsner, Daniel T. Gilbert, and Daniel L. Schacter. 2001. Do Amnesics Exhibit Cognitive Dissonance Reduction? The Role of Explicit Memory and Attention in Attitude Change. *Psychological Science* 12, 2 (2001), 135–140. http://www.jstor.org/stable/40063600

[57] Hannah Limerick, James W. Moore, and David Coyle. 2015. Empirical evidence for a diminished sense of agency in speech interfaces. *Conference on Human Factors in Computing Systems - Proceedings* 2015-April (2015), 3967–3970. Issue Figure 1. https://doi.org/10.1145/2702123.2702379 cognitive band; no co-actor(s).

[58] Qiuyu Lu, Jiawei Fang, Zhihao Yao, Yue Yang, Shiqing Lyu, Haipeng Mi, and Lining Yao. 2024. Enabling Generative Design Tools with LLM Agents for Building Novel Devices: A Case Study on Fluidic Computation Interfaces. *arXiv preprint arXiv:2405.17837* (2024).

[59] R Duncan Luce and Eugene Galanter. 1963. Psychophysical scaling. *Handbook of mathematical psychology* 1, 245-307 (1963), 50.

[60] Andreĭ Andreevitch Markov. 1906. Extension of the law of large numbers to quantities that depend on each other. *Proceedings of the Physics and Mathematics Society at Kazan University* 15, 2 (1906), 135–156.

[61] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. DirectGPT: A Direct Manipulation Interface to Interact with Large Language Models. *Conference on Human Factors in Computing Systems - Proceedings* (5 2024), 16. https://doi.org/10.1145/3613904.3642462/SUPPL_FILE/PN4616-SUPPLEMENTAL-MATERIAL-2.ZIP

[62] Justin Matejka, Michael Glueck, Tovi Grossman, and George Fitzmaurice. 2016. The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5421–5432. https://doi.org/10.1145/2858036.2858063

[63] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. 2021. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073* (2021).

[64] James W. Moore and Sukhvinder S. Obhi. 2012. Intentional binding and the sense of agency: A review. *Consciousness and Cognition* 21 (2012), 546–561. Issue 1. https://doi.org/10.1016/j.concog.2011.12.002

[65] James W. Moore, Daniel M. Wegner, and Patrick Haggard. 2009. Modulating the sense of agency with external cues. *Consciousness and Cognition* 18, 4 (Dec. 2009), 1056–1064. https://doi.org/10.1016/j.concog.2009.05.004

[66] Chong Mou, Xintao Wang, Jiechong Song, Ying Shan, and Jian Zhang. 2023. Dragondiffusion: Enabling drag-style manipulation on diffusion models. *arXiv preprint arXiv:2307.02421* (2023).

[67] Jun Nishida, Shunichi Kasahara, and Pedro Lopes. 2019. Demonstrating preemptive reaction: accelerating human reaction using electrical muscle stimulation without compromising agency. In *ACM SIGGRAPH 2019 Emerging Technologies* (Los Angeles, California) *(SIGGRAPH '19)*. Association for Computing Machinery, New York, NY, USA, Article 10, 2 pages. https://doi.org/10.1145/3305367.3327997

[68] Xingang Pan, Ayush Tewari, Thomas Leimkühler, Lingjie Liu, Abhimitra Meka, and Christian Theobalt. 2023. Drag Your GAN: Interactive Point-Based Manipulation on the Generative Image Manifold. In *ACM SIGGRAPH 2023 Conference Proceedings* (Los Angeles, CA, USA) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 78, 11 pages. https:

//doi.org/10.1145/3588432.3591500
[69] Mechthild Papoušek. 2004. *Regulationsstörungen der frühen Kindheit: Klinische Evidenz für ein neues diagnostisches Konzept.* na.
[70] Mechthild Papoušek. 2014. *Kommunikation und Sprachentwicklung im ersten Lebensjahr.* Springer Berlin Heidelberg, 69–80. https://doi.org/10.1007/978-3-642-39602-1_5
[71] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, and Dustin Tran. 2018. Image transformer. In *International conference on machine learning*. PMLR, 4055–4064.
[72] Claire Petitmengin. 2006. Describing one's subjective experience in the second person: An interview method for the science of consciousness. *Phenomenology and the Cognitive Sciences* 5, 3–4 (Nov. 2006), 229–269. https://doi.org/10.1007/s11097-006-9022-2
[73] Janet Rafner, Roger E. Beaty, James C. Kaufman, Todd Lubart, and Jacob Sherson. 2023. Creativity in the age of generative AI. *Nature Human Behaviour* 7, 11 (Nov. 2023), 1836–1838. https://doi.org/10.1038/s41562-023-01751-1
[74] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* 1, 2 (2022), 3.
[75] Ali Razavi, Aaron Van den Oord, and Oriol Vinyals. 2019. Generating diverse high-fidelity images with vq-vae-2. *Advances in neural information processing systems* 32 (2019).
[76] Daniel Roich, Ron Mokady, Amit H Bermano, and Daniel Cohen-Or. 2022. Pivotal tuning for latent-based editing of real images. *ACM Transactions on graphics (TOG)* 42, 1 (2022), 1–13.
[77] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.
[78] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. 2022. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems* 35 (2022), 36479–36494.
[79] Axel Sauer, Tero Karras, Samuli Laine, Andreas Geiger, and Timo Aila. 2023. StyleGAN-T: Unlocking the Power of GANs for Fast Large-Scale Text-to-Image Synthesis. *International Conference on Machine Learning* abs/2301.09515. https://arxiv.org/abs/2301.09515
[80] Axel Sauer, Katja Schwarz, and Andreas Geiger. 2022. Stylegan-xl: Scaling stylegan to large diverse datasets. In *ACM SIGGRAPH 2022 conference proceedings*. 1–10.
[81] Albrecht Schmidt, Passant Elagroudy, Fiona Draxler, Frauke Kreuter, and Robin Welsch. 2024. Simulating the human in HCD with ChatGPT: Redesigning interaction design with AI. *Interactions* 31, 1 (2024), 24–31.
[82] Yujun Shen and Bolei Zhou. 2021. Closed-form factorization of latent semantics in gans. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1532–1540.
[83] Roger N Shepard and Jacqueline Metzler. 1971. Mental rotation of three-dimensional objects. *Science* 171, 3972 (1971), 701–703.
[84] Jingyu Shi, Rahul Jain, Hyungjun Doh, Ryo Suzuki, and Karthik Ramani. 2023. An HCI-Centric Survey and Taxonomy of Human-Generative-AI Interactions. *arXiv preprint arXiv:2310.07127* (2023).
[85] Yujun Shi, Chuhui Xue, Jiachun Pan, Wenqing Zhang, Vincent YF Tan, and Song Bai. 2023. DragDiffusion: Harnessing Diffusion Models for Interactive Point-based Image Editing. *arXiv preprint arXiv:2306.14435* (2023).
[86] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*. PMLR, 2256–2265.
[87] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).
[88] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. 2020. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456* (2020).
[89] Matthis Synofzik, Gottfried Vosgerau, and Albert Newen. 2008. Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition* 17, 1 (2008), 219–239. https://doi.org/10.1016/j.concog.2007.03.010
[90] Daisuke Tajima, Jun Nishida, Pedro Lopes, and Shunichi Kasahara. 2022. Whose Touch is This?: Understanding the Agency Trade-Off Between User-Driven Touch vs. Computer-Driven Touch. *ACM Trans. Comput.-Hum. Interact.* 29, 3, Article 24 (jan 2022), 27 pages. https://doi.org/10.1145/3489608
[91] Zachary Teed and Jia Deng. 2020. Raft: Recurrent all-pairs field transforms for optical flow. In *ECCV*. Springer, 402–419.
[92] Ayush Tewari, Mohamed Elgharib, Gaurav Bharaj, Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zollhofer, and Christian Theobalt. 2020. Stylerig: Rigging stylegan for 3d control over portrait images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6142–6151.
[93] Aaron Van den Oord, Nal Kalchbrenner, Lasse Espeholt, Oriol Vinyals, Alex Graves, et al. 2016. Conditional image generation with pixelcnn decoders. *Advances in neural information processing systems* 29 (2016).
[94] Thomas Waltemate, Irene Senna, Felix Hülsmann, Marieke Rohde, Stefan Kopp, Marc Ernst, and Mario Botsch. 2016. The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* 02-04-November-2016 (11 2016), 27–35. https://doi.org/10.1145/2993369.2993381 cognitive band; no co-actor(s).
[95] Sheng-Yu Wang, David Bau, and Jun-Yan Zhu. 2022. Rewriting geometric rules of a GAN. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–16.
[96] Daniel M Wegner. 2003. The mind's best trick: How we experience conscious will. *Trends in cognitive sciences* 7, 2 (2003), 65–69.
[97] Daniel M Wegner. 2004. Précis of the illusion of conscious will. *Behavioral and Brain Sciences* 27, 5 (2004), 649–659.
[98] Daniel M Wegner and Thalia Wheatley. 1999. Apparent mental causation: Sources of the experience of will. *American psychologist* 54, 7 (1999), 480.
[99] Astrid Weiss. 2023. Human Interactions With (Embodid) AI - The Future of Authenticity in Human-AI Relations(hips). In *The De Gruyter Handbook of Robots in Society and Culture*, Autumn P. Edwards and Leopoldina Fortunati (Eds.). in press.
[100] Wen Wen and Hiroshi Imamizu. 2022. The sense of agency in perception, behaviour and human–machine interactions. *Nature Reviews Psychology* 1, 4 (April 2022), 211–222.
[101] Darrell M. West and John R. Allen. 2018. How artificial intelligence is transforming the world | Brookings — brookings.edu. https://www.brookings.edu/articles/how-artificial-intelligence-is-transforming-the-world/. [Accessed 03-04-2024].
[102] Debora Zanatto, Simone Bifani, and Jan Noyes. 2023. Constraining the Sense of Agency in Human-Machine Interaction. *International Journal of Human-Computer Interaction* (2023). https://doi.org/10.1080/10447318.2023.2189815 What-Whether-When paradigmDifferent ways of constraining the participants' actions. Also looks at time intervals between the action and outcome and found lower binding scores for longer intervals, which is in line with previous research as well.What and whether reduces IB, When does not. I miss the information on what method has been used for intentional binding? Libet clock or interval estimation? Because if interval estimation was used, the higher binding for longer intervals could be explained by more room to underestimate than for shorter intervals..
[103] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3836–3847.
[104] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.

A PRE-RENDERED IMAGES

Table 1: All images that have been pre-rendered indicating the model and the qualities on the outcome images. Each cell is one starting image described by [seed, x, y], indicating the seed within the model and the x and y values of the starting point on the image.

Lions						Humans					
stylegan2_lions_512_pytorch			stylegan2-afhqwild-512x512			stylegan2-ffhqu-1024x1024.pkl			stylegan2-ffhq-512x512.pkl		
Realism	**Distortion**	**Abstraction**	**Correlation**	**Distortion**	**Abstraction**	**Realism**	**Distortion**	**Abstraction**	**Correlation**	**Distortion**	**Abstraction**
53, 296, 143	16, 212, 269	36, 318, 271	35, 310, 198	367, 335, 292	19, 375, 250	7, 516, 461	511, 525, 456	550, 669, 559	5, 313, 272	361, 294, 250	397, 304, 257
82, 301, 237	518, 276, 258	638, 250, 213	42, 319, 314	377, 339, 327	30, 308, 219	9, 510, 514	515, 547, 571	568, 546, 530	22, 333, 238	373, 318, 268	406, 334, 264
337, 243, 367	557, 347, 346	957, 268, 190	51, 352, 210	417, 313, 240	40, 377, 305	24, 662, 569	520, 550, 563	653, 681, 432	60, 296, 253	379, 309, 259	413, 312, 288
475, 330, 226			247, 306, 241			32, 611, 573			63, 285, 254		
503, 293, 218			258, 366, 261			74, 488, 535			163, 289, 226		
777, 209, 205			365, 340, 233			472, 129, 925			352, 310, 250		