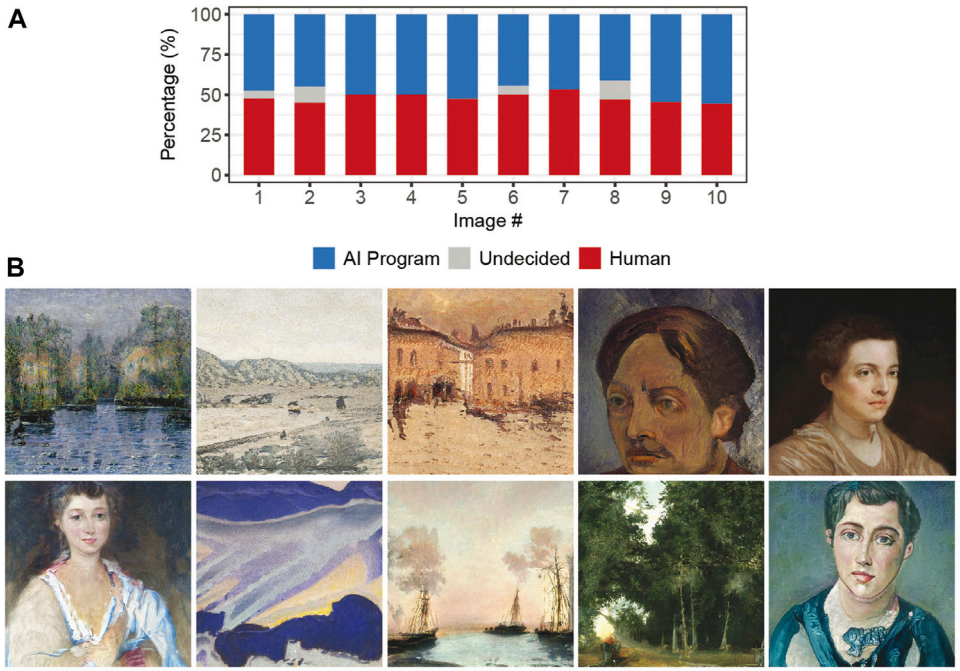

FIGURE 1 | Distribution of respondents' judgments of the top-10 images selected in our preliminary study for Studies 1 and 2 **(A)**. Selected images used in Studies 1 and 2 **(B)**.

the generated images to participants, who were asked to distinguish which images were generated by AI.

4.1 Methods

After agreeing to the research terms, study participants were told that they would be presented with a series of images generated by AI systems and human artists. Note that all images had been generated by the AI-generative model described above. Participants were instructed to indicate who they thought created each image—an AI program or a human artist. Participants were successively shown a random subset of 20 images in random order. Participants also had the option to indicate that they were unsure about its creator for each image. After evaluating all 20 images, participants were debriefed that an AI model had generated all images.

4.2 Participants

We recruited 45 respondents (22 men, 21 women, two others; 26 younger than 35 years old) through the Prolific crowdsourcing platform (https://www.prolific.co/; Palan and Schitter (2018)). Participants were required to have completed a minimum of 50 tasks in Prolific with at least a 95% approval rate. All respondents were United States nationals and were compensated $0.87 for the study.

4.3 Results

We chose images that were considered most ambiguous based on participants' ratings. This decision was made by the fact that GAN-based models are intentionally modeled to deceive a discriminator. These models' training process aims to teach a generator how to output ambiguous images that one cannot discriminate as either real (i.e., human-created) or artificial (i.e., AI-generated). Although another option would be to choose images that participants thought were human-created, we note that doing so could have made future participants suspect the images' origin. Hence, to mitigate possible deception effects, we decided to discard images that were perceived to have been created by human artists.

None of the images had a majority of respondents being unsure about its provenance. We thus used Shannon Entropy to compute image ambiguity across responses indicating that humans or AI systems created the images. We selected the top-10 images in terms of ambiguity and used them for all subsequent studies. Of the ten images, five are landscapes, four are portraits, and one is an abstraction. Qualitative analysis of all 58 images showed that more realistic images were often perceived as human-created. On the other hand, abstractions were more frequently viewed as AI-generated. **Figure 1A** presents the distribution of responses for the selected images, and **Figure 1B** shows them. All images are made available in the study's online repository for future research.

5 STUDY 1

Study 1 examined whether Gunkel's social-relational approach to electronic agents' moral standing could be applied to the context of AI-generative art. Our study employed a between-subjects design where participants interacted with AI-generated images

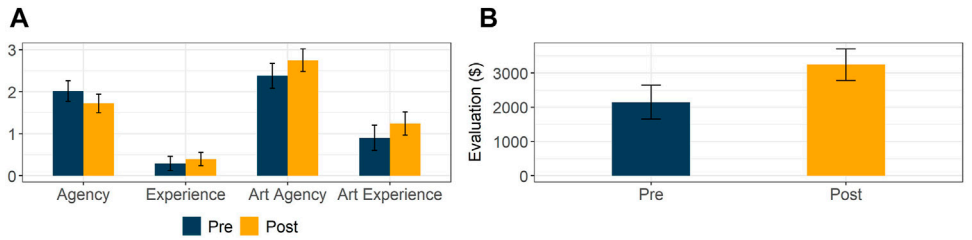

FIGURE 2 | Attribution of agency, experience, art agency, and art experience to an AI system before and after being exposed to AI-generated paintings in Study 1 **(A)**. Marginal mean evaluation across all ten images depending on treatment group **(B)**.

similar results when accounting for respondents' attitudes towards AI and their previous knowledge of computer science and art-related subjects.

5.4 Discussion

Whether participants interacted with AI-generated images before or after attributing moral agency and patiency to the system did not influence its perceived moral standing. We observed a significant difference in participants' perception of the AI system's capacity to create and experience art depending on the treatment condition. This effect, however, disappeared once we controlled for participants' attitudes towards the AI systems' outputs, i.e., the average price assigned to AI-generated art. It may well be the case that our proposed interaction with AI-generated art is not as strong a stimuli as the significant social interactions that authors defend to be crucial components of moral standing.

Nevertheless, study participants ascribed the ability to create art to the AI system although it was not described as an "artist," nor their outputs were introduced as "art." This specific artistic notion of the agency was perceived as more significant to the AI-generative system than the more general conception of agency captured by the mind perception questionnaire. In a similar vein, our results indicate that AI systems were attributed some ability to experience art even though they were not perceived to have the experience dimension of mind.

Finally, we observed a significant difference across treatment groups by expanding our analysis to how participants responded to AI-generated paintings. Even after controlling for individual variations through a mixed-effects model, AI-generated images were valued lower by participants who attributed moral standing to the AI system before interacting with its images. This result suggests that nudging participants to think about an AI system's mind (e.g., its agency and patiency) could negatively influence how much they value its outputs. That is, the act of evaluating an AI system's moral status could influence how people interact with them.

6 STUDY 2

Study 2 inquired whether Coeckelbergh's socio-relational approach to electronic agents' indirect moral status could be extended to the context of AI-generative art. The author suggests that electronic agents could be granted moral standing if others have a valuable relationship with them, i.e., one should respect these systems' interests due to their extrinsic value. Hence, our study was designed to randomly assign participants to treatment groups that show how others perceived AI-generated images, e.g., by under- or overvaluing them.

6.1 Methods

After agreeing to the research terms, participants were told that some existing AI systems could generate images and that they would be shown some examples throughout the study. Each participant was randomly assigned to one of four treatment groups. Those assigned to the *pre* condition took part in a study similar to the *pre* condition in Study 1, i.e., they attributed moral status before interacting with a series of AI-generated images. Participants allocated to the *undervalue*, *median*, and *overvalue* conditions were presented a study design similar to Study 1's *post* condition, where participants first evaluated a set of AI-generated paintings and then answered questions concerning their creator's moral status.

Study 2 differed from the previous study in that participants were shown additional information during the art evaluation step. After evaluating each of the images, participants were shown how other respondents evaluated the same painting depending on the treatment condition they were assigned to. They were subsequently asked to modify their initial evaluation if they desired to do so. Participants assigned to *pre* and *median* conditions were shown median values calculated from Study 1's responses.[1] Those in the *undervalue* and *overvalue* groups were presented to evaluations three times lower or larger than those presented in the other two conditions. This design choice aimed to elucidate the AI system's extrinsic value, which Coeckelbergh argues to be crucial for electronic agents' moral standing.

All participants responded to the same mind perception questionnaire and art-related questions from Study 1. We additionally asked participants to rate the AI-generative system's moral standing concerning six statements. Respondents

[1]Due to a programming error, median values were calculated with respect to the order images were shown to participants in Study 1. For instance, image #1's median value was determined by the median evaluation of the first image shown to each participant. Note that the image order was randomized between participants. Our study conditions should not be affected by this error, i.e., all images were overvalued or undervalued on their respective treatment conditions.

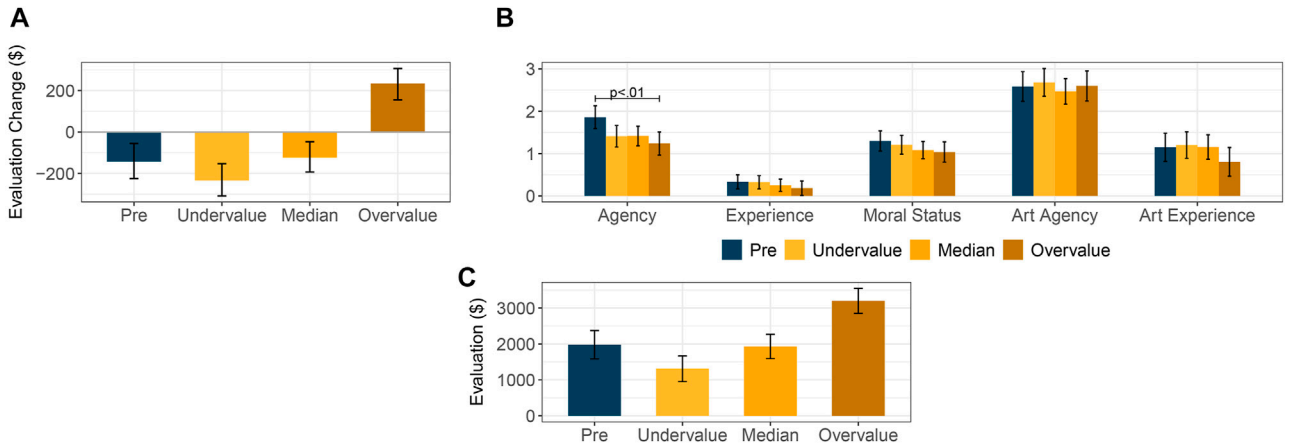

FIGURE 3 | To what extent participants modified their initial art evaluation after treatment in Study 2 **(A)**. Attributions of agency, experience, moral status, art agency, and art experience to the AI system depending on the condition participants were assigned to in Study 2 **(B)**. Marginal mean evaluation across all ten images depending on treatment group **(C)**.

We expanded Study 2 to include a novel measure of perceived moral standing independent of an entity's perceived experience covered by the mind perception questionnaire. This was done because the social-relational approach to electronic agents' moral standing challenges perspectives that defend experience-related capacities as preconditions for moral status. Nevertheless, we did not find any significant difference between treatment conditions in both attributions of experience and our proposed moral standing measure. These results corroborate our findings from Study 1 by showing that interacting with AI-generated outputs should not influence people's ascription of moral standing.

Nudging people to think about the mind of an AI system did not necessarily influence how they valued AI-generated art in Study 2. Our results instead suggest that overvaluing AI-generated art could influence how people perceive it. We hypothesize that the treatment conditions' social influence mitigated any possible effect of considerations about an AI system's mind similar to those found in Study 1. Similar to how past auctions of AI-generated art were presented to the public (Cohn, 2018; Ives, 2021), overvaluing these outputs could influence how much people value them.

7 GENERAL DISCUSSION

Inspired by Gunkel's and Coeckelbergh's social-relational approaches to robots' moral standing, we conducted two studies to understand whether a similar perspective would influence people's ascription of moral status to a nonsocial automated agent, namely an AI-generative system. We first identified a set of ten AI-generated images that were used in subsequent studies. Study 1 inquired whether interacting with these images would influence people's ascription of moral agency and patiency to their creator—as suggested by Gunkel (2018b). Study 2 asked whether highlighting an AI system's extrinsic value by undervaluing or overvaluing its images affected participants' attribution of agency, experience, and moral status, as proposed by Coeckelbergh (2020b). The current research took a novel experimental approach to the normative debate of robot rights in the context of AI-generated art.

We employed a series of measures to quantify AI systems' perceived moral (and artistic) standing. Interacting with AI-generated art did not significantly impact how participants perceived the system's ability to create art, experience art, and the experience dimension of mind in both Studies 1 and 2. The latter was measured by a mind perception questionnaire, whose measure has been shown to correlate with the recognition of moral rights (Waytz et al., 2010; Gray et al., 2007). Study 2 also showed that interacting with AI-generated art did not influence the AI system's perceived moral standing in a novel measure of moral consideration independent of the system's experience.

Study 2's participants attributed lower levels of agency to AI systems after interacting with overvalued AI-generated art. This finding suggests that seeing others overvaluing AI systems' abilities could negatively influence their perceived agency. This finding may be contrary to what one would expect. Similar to Coeckelbergh's approach to AI systems' patiency, highlighting the system's creative value by overvaluing its generated images should, at first thought, increase their perceived (artistic) agency.

Finally, Study 1 suggests that nudging participants to think about an AI systems' mind could lead to a lower appreciation of AI-generated art. A possible interpretation is that machine creativity is not valued to the same extent as its human counterparts, particularly when AI systems' lack of humanness and mind becomes apparent. As argued by some scholars, AI-generated art may lack the meaning necessary to be considered art—such meaning can only emerge from human artistic communication (Elgammal, 2020). Another possible explanation is that art is also evaluated by the effort put into its creation. More realistic images in our Experimental Setting were often attributed to human artists, while abstractions were usually viewed as AI-generated. Participants might have judged