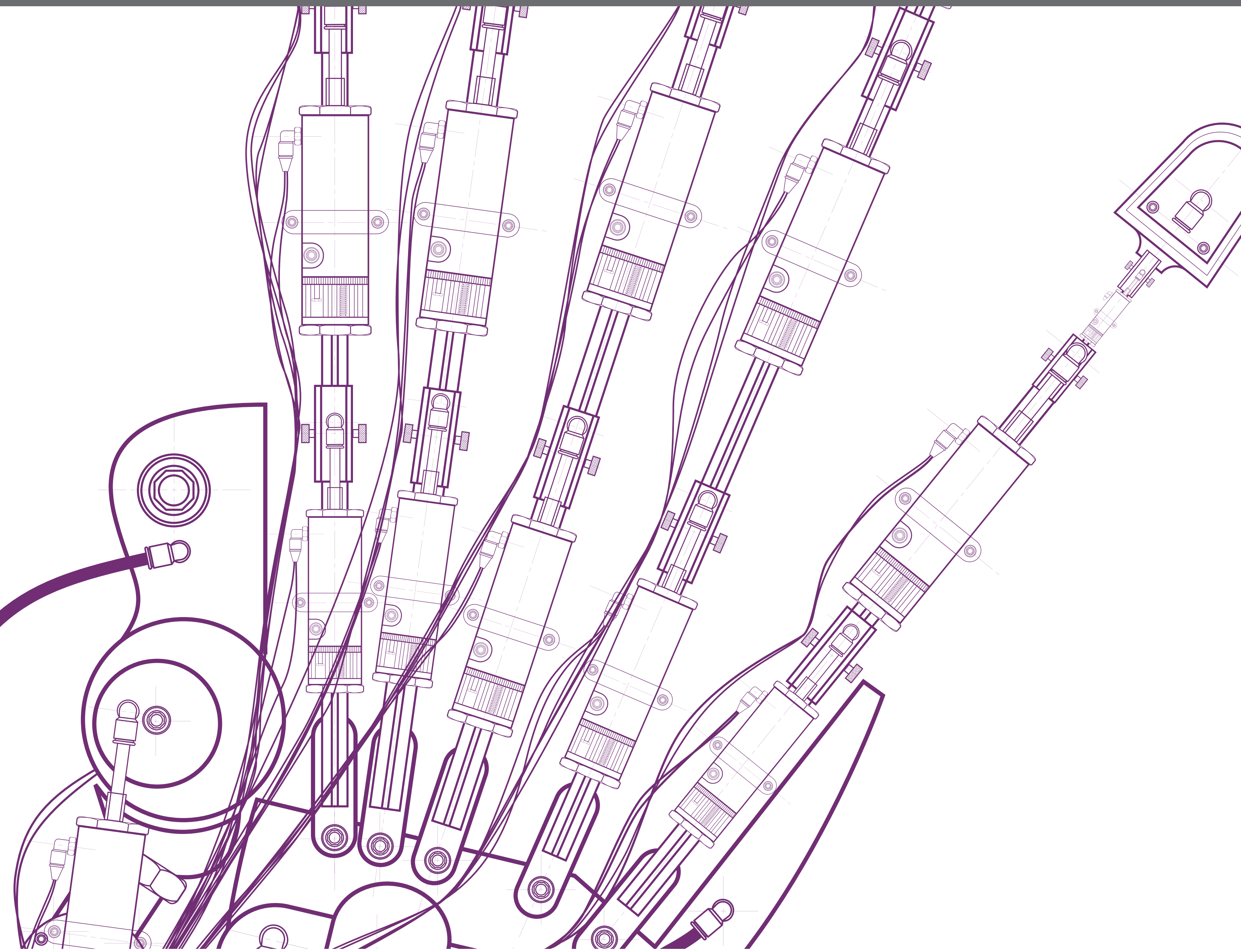
CREATIVITY AND ROBOTICS

EDITED BY: Patricia Alves-Oliveira, Amy LaViers, Peter H. Kahn,
Goren Gordon, Maya Cakmak and Vasanth Sarathy
PUBLISHED IN: Frontiers in Robotics and AI



frontiers Research Topics

Frontiers eBook Copyright Statement

ISSN 1664-8714
ISBN 978-2-83250-384-3
DOI 10.3389/978-2-83250-384-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

CREATIVITY AND ROBOTICS

Topic Editors:
Patricia Alves-Oliveira, University of Washington, United States
Amy LaViers, The RAD Lab, United States
Peter H. Kahn, University of Washington, United States
Goren Gordon, Tel Aviv University, Israel
Maya Cakmak, University of Washington, United States
Vasanth Sarathy, Tufts University, United States

Citation: Alves-Oliveira, P., LaViers, A., Kahn, P. H., Gordon, G., Cakmak, M., Sarathy, V., eds. (2022). Creativity and Robotics. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-384-3

Table of Contents

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 23 December 2020
doi: 10.3389/frobt.2020.592382

Feature Guided Search for Creative Problem Solving Through Tool Construction

Lakshmi Nair* and Sonia Chernova

Georgia Institute of Technology, Atlanta, GA, United States

Robots in the real world should be able to adapt to unforeseen circumstances. Particularly in the context of tool use, robots may not have access to the tools they need for completing a task. In this paper, we focus on the problem of tool construction in the context of task planning. We seek to enable robots to construct replacements for missing tools using available objects, in order to complete the given task. We introduce the Feature Guided Search (FGS) algorithm that enables the application of existing heuristic search approaches in the context of task planning, to perform tool construction efficiently. FGS accounts for physical attributes of objects (e.g., shape, material) during the search for a valid task plan. Our results demonstrate that FGS significantly reduces the search effort over standard heuristic search approaches by $\approx 93\%$ for tool construction.

Keywords: tool construction, creative problem solving, task planning, heuristic search, adaptive robotic systems

OPEN ACCESS

Edited by:
Vasanth Sarathy,
Tufts University, United States

Reviewed by:
Safinah Ali,
Massachusetts Institute of Technology, United States
Lorenzo Jamone,
Queen Mary University of London, United Kingdom

****Correspondence:***
Lakshmi Nair
lnair3@gatech.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *06 August 2020*
Accepted: *01 December 2020*
Published: *23 December 2020*

Citation:
Nair L and Chernova S (2020) Feature Guided Search for Creative Problem Solving Through Tool Construction. Front. Robot. AI 7:592382. doi: 10.3389/frobt.2020.592382

1. INTRODUCTION

Humans often show remarkable improvisation capabilities, particularly in times of crises. The makeshift carbon dioxide filter constructed on board the Apollo 13 (Cass, 2005), and the jury-rigged ventilators built to combat equipment shortages during COVID-19 (Turner et al., 2020), are examples of human ingenuity in the face of uncertainty. In addition to humans, other primates and certain species of birds have also been shown to creatively accomplish tasks by constructing tools from available objects, such as sticks and stones (Jones and Kamil, 1973; Toth et al., 1993; Stout, 2011). While the capability to construct tools is often regarded as a hallmark of sophisticated intelligence, similar improvisation capabilities are currently beyond the scope of existing robotic systems. The ability to improvise and construct necessary tools can greatly increase robot adaptability to unforeseen circumstances, enabling robots to handle any uncertainties or equipment failures that may arise (Atkeson et al., 2018).

In this paper, we focus on the problem of tool construction in the context of task planning. Specifically, we address the scenario in which a robot is provided with a task that requires certain tools that are missing or unavailable. The robot must then derive a task plan that involves constructing an appropriate replacement tool from objects that are available to it, and use the constructed tool to accomplish the task. Existing work that addresses the problem of planning in the case of missing tools focuses on directly substituting the missing tool with available objects (Agostini et al., 2015; Boteanu et al., 2015; Nyga et al., 2018). In contrast, this is the first work to address the problem through the construction of replacement tools, by introducing a novel approach called Feature Guided Search (FGS). FGS enables efficient application of existing heuristic search algorithms in the context of task planning in order to perform tool construction by accounting for physical attributes of objects (e.g., shape, material) during the search for a valid task plan.

Heuristic search algorithms, such as A^* and enforced hill-climbing (EHC), have been successfully applied to planning problems in conjunction with heuristics such as cost-optimal landmarks (Karpas and Domshlak, 2009) and fast-forward (Hoffmann and Nebel, 2001), respectively. However, the application of heuristic search algorithms to perform tool construction in the context of task planning can be challenging. For example, consider a task where the goal of the robot is to hang a painting on the wall. In the absence of a hammer that is required for hammering a nail to complete the task, the robot may choose to construct a replacement for the hammer using the objects available to it. How does the robot know which objects should be combined to construct the replacement tool? One possible solution is for the user to manually encode the correct object combination in the goal definition, and the search procedure would find it. However, it is impractical for the user to know and encode the correct object combination to use, for all the objects that the robot could possibly encounter. Alternatively, the robot can autonomously attempt every possible object combination until it finds an appropriate tool construction for completing the task. However, this would require a prohibitive number of tool construction attempts. Further, what if the robot *cannot* construct a good replacement for a hammer using the available objects, but can instead construct a makeshift screwdriver to tighten a screw and complete the task? In this case, the task plan would also have to be adapted to appropriately use the constructed tool, i.e., "tighten" a screw with the screwdriver instead of "hammering" the nail. In order to address these challenges, FGS combines existing planning heuristics with a score that is computed from input point clouds of objects indicating the best object combination to use for constructing a replacement tool. The chosen replacement tool then in turn guides the correct action(s) to be executed for completing the task (e.g., "tighten" vs. "hammering"). Hence, our algorithm seeks to: (a) eliminate the need for the user to specify the correct object combination, thus enabling the robot to autonomously choose the right tool construction based on the available objects and the task goal, (b) minimize the number of failed tool construction attempts in finding the correct solution, and (c) adapt the task plan to appropriately use the constructed replacement tool.

Prior work by Nair et al. introduced a novel computational framework for performing tool construction, in which the approach takes an input action, e.g., "flip," in order to output a ranking of different object combinations for constructing a tool that can perform the specified action, e.g., constructing a spatula (Nair et al., 2019a,b). For performing the ranking, the approach scored object combinations based on the shape and material properties of the objects, and whether the objects could be attached appropriately to construct the desired tool. In contrast, this work focuses on the application of heuristic search algorithms such as A^*, to the problem of tool construction in the context of task planning. In this case, the robot is provided an input *task*, e.g., "make pancakes," that requires tools that are inaccessible to the robot, e.g., a missing spatula. The robot must then output a *task plan* for making pancakes, that involves constructing an appropriate replacement tool from available objects, and adapting the task plan to use the constructed tool for completing the task. Thus, prior work takes an action as input, and outputs a ranking of object combinations. In contrast, our work takes a task as input, and outputs a task plan that involves constructing and using an appropriate replacement tool. Hence, our work relaxes a key assumption of the prior work that requires the input action to be specified. Our approach directly uses the score computation methodology described in prior work (Nair et al., 2019a,b), but combines it with planning heuristics to integrate tool construction within a task planning framework.

Our core contributions in this paper include:

- Introducing the Feature Guided Search (FGS) approach that integrates reasoning about physical attributes of objects with existing heuristic search algorithms for efficiently performing tool construction in the context of task planning.
- Improving upon prior work by enabling the robot to automatically choose the correct tool construction and the appropriate action based on the task and available objects, thus eliminating the need to explicitly specify an input action as assumed in prior work.

We evaluate our approach in comparison to standard heuristic search baselines, on the construction of six different tool types (hammer, screwdriver, ladle, spatula, rake, and squeegee), in three task domains (wood-working, cooking, and cleaning). Our results show that FGS outperforms the baselines by significantly reducing computational effort in terms of number of failed construction attempts. We also demonstrate the adaptability of the task plans generated by FGS based on the objects available in the environment, in terms of executing the correct action with the constructed tool.

2. RELATED WORK

Prior work by Sarathy and Scheutz have focused on formalizing creative problem solving in the context of planning problems (Sarathy and Scheutz, 2017, 2018). They define the notion of "Macgyver-esque" creativity as embodied agents that can "*generate, execute, and learn strategies for identifying and solving seemingly unsolvable real-world problems*" (Sarathy and Scheutz, 2017). They formalize Macgyvering problems (MGP) with respect to an agent t, as a planning problem in the agent's world $\mathbb{W}_t$, that has a goal state g currently unreachable by the agent. As described in their work, solving an MGP requires a domain extension or contraction through perceiving the agent's environment and self. Prior work by Sarathy also provide an in-depth discussion of the cognitive processes involved in creative problem solving in detail, by leveraging existing work in Neuroscience (Sarathy, 2018). Prior work by Olteţeanu and Falomir has also looked at the problem of Object Replacement and Object Composition (OROC) situated within a cognitive framework called, the Creative Cognitive Framework (CreaCogs) (Olteţeanu and Falomir, 2016). Their work utilizes a knowledge base that semantically encodes object properties and relationships in order to reason about alternative uses for objects to creatively solve problems. The semantic relationships themselves are currently encoded a-priori. Similar work by

Freedman et al. has focused on the integration of analogical reasoning and automated planning for creative problem solving by leveraging semantic relationships between objects (Freedman et al., 2020). They present the Creative Problem Solver (CPS), that uses large-scale knowledge bases to reason about alternate uses of objects for creative problem solving. In contrast to reasoning about objects, prior work by Gizzi et al. has looked at the problem of discovering new actions for creative problem solving, enabling the robot to identify previously unknown actions (Gizzi et al., 2019). Their work applies action segmentation and change-point detection to previously known actions to enable a robot to discover new actions. The authors then apply breadth-first search and depth-first search in order to derive planning solutions using the newly discovered actions.

In related work, Erdogan and Stilman (2013) described techniques for Automated Design of Functional Structures (ADFS), involving construction of navigational structures, e.g., stairs or bridges. They introduce a framework for effectively partitioning the solution space by inducing constraints on the design of the structures. Further, Tosun et al. (2018) have looked at planning for construction of functional structures by modular robots, focusing on identifying features that enable environment modification in order to make the terrain traversable. In similar work, Saboia et al. (2018) have looked at modification of unstructured environments using objects, to create ramps that enhance navigability. More recently, Choi et al. (2018) extended the cognitive architecture ICARUS to support the creation and use of functional structures such as ramps, in abstract planning scenarios. Their work focuses on using physical attributes of objects that is encoded a-priori, such as weight and size, in order to reason about the construction and stability of navigational structures. More broadly, these approaches are primarily focused on improving robot navigation through environment modification as opposed to construction of tools. Some existing research has also explored the construction of simple machines such as levers, using environmental objects (Levihn and Stilman, 2014; Stilman et al., 2014). Their work formulates the construction of simple machines as a constraint satisfaction problem where the constraints represent the relationships between the design components. The constraints in their work limit the variability of the simple machines that can be constructed, focusing only on the placement of components relative to one another, e.g., placing a plank over a stone to create a lever. Additionally, Wicaksono and Sheh (2017) have focused on using 3D printing to fabricate tools from polymers. Their work encodes the geometries of specific sub-parts of tools, and enables the robot to experiment with different configurations of the fabricated tools to evaluate their success for accomplishing a task.

The work described in this paper differs from the research described above in that we focus on creative problem solving through tool construction. Specifically, we focus on planning tasks in which the required tools need to be constructed from available objects. Two key aspects of our work that further distinguish it from existing research include: (i) sensing and reasoning about physical features of objects, such as shape, material, and the different ways in which objects can be attached, and (ii) improving the performance of heuristic search algorithms for tool construction in the context of task planning, by incorporating the physical properties of objects during the search for a task plan.

3. APPROACH

In this section, we begin by discussing some background details regarding heuristic search, followed by specific implementation details of FGS.

3.1. Heuristic Search

Heuristic search algorithms are guided by a cost function $f(s) = g(s) + h(s)$, where $g(s)$ is the best-known distance from the initial state to the state s, and $h(s)$ is a heuristic function that estimates the cost from s to the goal state. An admissible heuristic never overestimates the path cost from any state s to the goal (Hart et al., 1968; Zhang et al., 2009). A consistent heuristic holds the additional property that, if there is a path from a state x to a state y, then $h(x) \leq d(x, y) + h(y)$, where $d(x, y)$ is the distance from x to y (Hart et al., 1968). Most heuristic search algorithms, including A^*, operate by maintaining a priority queue of states to be expanded (the open list), sorted based on the cost function. At each step, the state with the least cost is chosen, expanded, and the successors are added to the open list. If a successor state is already visited, the search algorithm may choose to re-expand the state, *only if* the new path cost to the state is lesser than the previously found path cost (Bagchi and Mahanti, 1983). The search continues until the goal state is found, or the open list becomes empty, in which case no plan is returned.

3.2. Feature Guided Search

We now describe the implementation of FGS[1]. For the purposes of this explanation, we present our work in the context of A^*, though our approach can be easily extended to other heuristic search algorithms as demonstrated in our experiments. Let S denote the set of states, A denote the set of actions, γ denote state transitions, s_i denote the initial state, and s_g denote the goal state. For the planning task, we consider the problem to be specified in Planning Domain Definition Language (PDDL) (McDermott et al., 1998), consisting of a domain definition $\mathcal{P}_{\mathcal{D}} = (S, A, \gamma)$, and a problem/task definition $\mathcal{P}_{\mathcal{T}} = (\mathcal{P}_{\mathcal{D}}, s_i, s_g)$. Further, we use O to denote a set of n objects in the environment available for tool construction, $O = \{o_1, o_2, ...o_n\}$.

Since our work focuses on tools, we assume that some action(s) in A are parameterized by a set of object(s) $O_a \subseteq O$, that are used to perform the action. Specifically for tool construction, we explicitly define an action "join(O_a)", where $O_a = \{o_1, o_2, ...o_m\}, m \leq n$, parameterized by objects that can be joined to construct a tool for completing the task. For example, the action "join-hammer(O_a)" allows the robot to construct a hammer using the objects O_a that parameterize the action. For actions that are not parameterized by any object, $O_a = \emptyset$. Our approach seeks to assign a "feature score" to the objects in O_a,

[1] All source code including problem and domain definitions, are publicly available at: https://github.com/Lnair1993/Tool_Macgyvering.

Algorithm 1: Feature guided A^* search.

1 **Function** `Search(`$\mathcal{P}_\mathcal{D}, \mathcal{P}_\mathcal{T}$, *trust=true*`):`
2 $s_i, s_g = \text{extractStates}(\mathcal{P}_\mathcal{T})$
3 $A = \text{extractActions}(\mathcal{P}_\mathcal{D})$
4 $O = \text{extractObjects}()$
5 $O_{reject} = [\ ]$
6 openList = [ ]
7 setPathCost(s_i, 0) *// Set initial state's g(s) and f(s) to 0*
8 openList. `add`(s_i, 0)
9 **while** *OpenList* ***not*** *empty* **do**
10 currState = $argmin_s(f(s))\ \forall\ s \in$ openList
11 openList. `pop`(currState)
12 **if** *currState* = s_g **then**
13 **return** extractPlan(s_g, s_i)
14 nextStates = getNext(currState, A)
15 **for** $(s, a, O_a) \in$ *nextStates* **do**
16 $g(s)$ = computePathCost(s, currState) *// Get current path cost*
17 $c(s)$ = getPathCost(s) *// Get previous best known path cost*
18 **if** $g(s) \geq c(s)$ **then**
19 **continue**
20 **else**
21 setPathCost($s, g(s)$) *// Update lower costs as new paths are found*
22 **end**
23 $h(s)$ = computeHeuristic(s)
24 $\phi(s)$ = **featureScore**(s, a, O_a, *trust*) *// Compute the feature score: Algorithm 2*
25 **if** $\phi(s) = -\infty$ **then**
26 O_{reject}. `add`(O_a, a) *// Track rejected combinations*
27 $f(s) = g(s) + h(s) - \phi(s)$
28 **if** $f(s) = \infty$ **then**
29 **continue**
30 openList. `add`($s, f(s)$)
31 **end**
32 **end**
33 **if** O_{reject} *not* $\emptyset$ **then**
34 `Search(`$\mathcal{P}_\mathcal{D}, \mathcal{P}_\mathcal{T}$, *trust = false*`)` *// Re-attempt without trusting all sensors*
35 **return** $\emptyset$ *// No plan found*

indicating their fitness for performing the action a. Thus, given different sets of objects O_a that are valid parameterizations of a, the feature score can help guide the search to generate task plans that involve using the objects that are most appropriate for performing the action. In the context of tool construction, the feature score guides the search to generate task plans that involve joining the most appropriate objects for constructing the replacement tool, given the objects available in the environment. Feature scoring can also potentially reject objects that are unfit for tool construction.

Our approach is presented in Algorithm 1. The search algorithm extracts information regarding the initial state s_i and goal state s_g from the task definition (Line 2). The set of actions A is extracted from the domain definition $\mathcal{P}_\mathcal{D}$ (Line 3). The agent extracts the objects in its environment from an RGB-D observation of the scene through point cloud segmentation and clustering (Line 4). We initialize the open list (*openList*) as a priority queue with the initial state s_i and cost of 0 (Lines 6–8). Lines 9–32 proceed according to the standard A^* search algorithm, except for the computation of the feature score in Line 24. While the open list is not empty, we select the state with the lowest cost function (Line 10,11). If the goal is found, the plan is extracted (Lines 12–13), otherwise the successor states are generated (Line 14). For each successor state s, the algorithm computes the path cost $g(s)$ from the current state *currState* to s (Line 16). The algorithm then retrieves the best known path cost $c(s)$ for the state from its previous encounters (Line 17). If the state was not previously seen, $c(s) = \infty$. In Lines 18–22, the algorithm compares the best known path cost to the current path cost, and updates the best known path cost if $g(s) < c(s)$. The algorithm then computes the heuristic $h(s)$ (Line 23), and the feature score $\phi(s)$ (Line 24). The algorithm also maintains a list of object combinations that were rejected by feature scoring (i.e., assigned a score of $-\infty$), in O_{reject} (Line 26). The final cost is computed as $f(s) = g(s) + h(s) - \phi(s)$ (Line 27; We expand more on our choice of cost function in section 3.5). If $f(s) \neq \infty$, then the state is added to the open list, prioritized by the cost. The search continues until a plan is found, or exits if the open list becomes empty. If no plan was found, the search is reattempted (Line 34) by modifying the feature score computation (described in section 3.3). If all search attempts fail, the planner returns a failure with no plan found. In the following section we discuss the computation of the feature score in detail.

3.3. Feature Score Computation

In this section, we describe the computation of the feature score for a given set of objects O_a that parameterize an action a. Note that, in this work the feature score computation focuses on the problem of tool construction. However, FGS can potentially be extended to other problems such as tool substitution, by computing similar feature scores as described in prior work (Abelha et al., 2016; Shrivatsav et al., 2019). Given n objects, tool construction presents a challenging combinatorial problem with a state space of size ${}^nP_m$, assuming that we wish to construct a tool with $m \leq n$ objects. Thus, $O_a = \{o_1, o_2, ...o_m\}$ denotes a specific permutation of m objects. Inspired by existing tool-making studies in humans (Beck et al., 2011), prior work introduced a multi-objective function for evaluating the fitness of objects for tool construction (Nair et al., 2019b), that we apply in this work for feature scoring. The proposed multi-objective function included three considerations: (a) shape fitness of the objects for performing the action, (b) material fitness of the objects for performing the action, and (c) evaluating whether the objects in O_a can be attached to construct the tool.

The calculation of each of the three metrics above relies on real-world sensing, which can be noisy. This can result in false negative predictions, that eliminate potentially valid object combinations from consideration. In particular, prior work has shown that false negatives in material and attachment predictions have caused $\approx$ 4% of tool constructions to fail (Nair et al., 2019b). To address the problem of false negatives in material and attachment predictions, we introduce the notion of "sensor trust" in this work. Prior work that has looked at accounting for sensor trust has introduced the notion of "trust weighting" to use continuous values to appropriately weigh the sensor inputs (Pierson and Schwager, 2016). In contrast, the sensor trust parameter in our work is a *binary value* that determines whether the material and attachment predictions should be believed by the robot and included in the feature score computation. This is because material and attachment scores are hard constraints and not continuous, i.e., they are $-\infty$ for objects that are not suited for tool construction (we describe this further in later sections). Hence, a continuous weighting on the material and attachment scores is not appropriate for our work.

Our feature score computation approach is described in Algorithm 2. For actions that are not parameterized by objects, the approach returns 0 (Lines 2-3). If the trust parameter is set to *true*, the feature score computation incorporates shape, material, and attachment predictions. (Lines 5–12 of Algorithm 2; section 3.4.1 for details). If the trust parameter is set to *false*, the feature score computation only includes shape scoring (Lines 14–19 of Algorithm 2; section 3.4.2 for details). Thus, we describe two modes of feature score computation that is influenced by the sensor trust parameter. In the following sections, we briefly describe the computation of shape, material and attachment predictions, and for a more detailed implementation of each method, we refer the reader to Nair et al. (2019b) and Shrivatsav et al. (2019).

Algorithm 2: Feature score Computation

1 **Function** `FeatureScore`($s, a, O_a, trust = true$):
2 **if** O_a ***is*** *empty* **then**
3 **return** 0
4 **if** *trust* **then**
5 **if** *canAttach*(O_a, a) **then**
6 $\phi^s_{shape}(O_a) = \text{ShapeFit}(O_a, a)$ *// Sensors are fully trusted - section 3.4.1*
7 $\phi^s_{mat}(O_a) = \text{MaterialFit}(O_a, a)$
8 $\phi(s) = \lambda_1 * \phi^s_{shape}(O_a) + \lambda_2 * \phi^s_{mat}(O_a)$ *// The weighted sum is assigned to s*
9 **return** $\phi(s)$
10 **else**
11 **return** $-\infty$
12 **end**
13 **else**
14 **if** $(O_a, a) \in O_{reject}$ **then**
15 $\phi^s_{shape}(O_a) = \text{ShapeFit}(O_a, a)$ *// Not fully trust sensors - section 3.4.2*
16 **return** $\phi^s_{shape}(O_a)$ *// Evaluate objects that were previously rejected*
17 **else**
18 **return** $-\infty$
19 **end**
20 **end**

3.3.1. Shape Scoring

Shape scoring seeks to predict the shape fitness of the objects in O_a for performing the action a. This is indicated by the *ShapeFit*() function in Algorithm 2. In this work, we consider tools to have action parts and grasp parts[2]. Thus, $m = 2$ and the set of objects O_a consists of two objects, i.e., $|O_a| = 2$. Further, the ordering of objects in O_a indicates the correspondence of the objects to the action and grasp parts.

For shape scoring, we seek to train models that can predict whether an input point cloud is suited for performing a specific action. We formulate this as a binary classification problem. We represent the shape of the input object point clouds using Ensemble of Shape Functions (ESF) (Wohlkinger and Vincze, 2011) which is a 640-D vector, shown to perform well in representing object shapes for tools (Schoeler and Wörgötter, 2015; Nair et al., 2019b). We then train independent neural networks that take an input ESF feature, and output a binary label indicating whether the input shape feature is suited for performing a specific action. Thus, we train separate neural networks, one for each action[3]. More specifically, we train separate networks for the tools' action parts, e.g., the head of a hammer or the flat head of a spatula, and for a supporting function: "*Handle*," which refers to the tools' grasp part, e.g., hammer handle.

For the score prediction, given a set of objects O_a to be used for constructing the tool, let $\mathcal{K}$ denote the set of objects in O_a that are candidates for the action parts of the final tool, and let $O_a - \mathcal{K}$ be the set of candidate grasp parts. Then the shape score $\phi^s_{shape}(O_a)$ is computed by using the trained networks as:

$$\phi^s_{shape}(O_a) = \prod_{o_i \in \mathcal{K}} p(action|o_i) \prod_{o_i \in O_a - \mathcal{K}} p(handle|o_i) \qquad (1)$$

Where, p is the prediction confidence of the corresponding network. Thus, we combine prediction confidences for all action parts and grasp parts. For example, for the action "join-hammer(O_a)" where O_a consists of two objects (o_1, o_2), the shape score $\phi^s_{shape}(O_a) = p(hammer\_head|o_1) * p(handle|o_2)$.

3.3.2. Material Scoring

Material scoring seeks to predict the material fitness of the objects in O_a for performing the action a. This is indicated by the *MaterialFit*() function in Algorithm 2. In this work, we make three simplifying assumptions. Firstly, we consider the construction of rigid tools which covers a vast range of real-world examples (Myers et al., 2015; Abelha et al., 2016). Secondly, we consider the material properties of the action parts of the

[2] As in prior work, this covers the vast majority of tools (Myers et al., 2015; Abelha and Guerin, 2017).

[3] The advantage of the binary classification is that for new actions, additional networks can be trained independently without affecting other networks.

TABLE 1 | Table indicating appropriate materials for action parts of different tools.

Tool	Material (Action part)
Hammer	Metal, Wood
Screwdriver	Plastic, Metal
Ladle	Plastic, Wood, Metal
Spatula	Plastic, Wood, Metal
Rake	Plastic, Wood, Metal
Squeegee	Foam

tool since the action parts are more critical to performing the action with the tool (Shrivatsav et al., 2019). Lastly, we assume that the materials that are appropriate for different tools is provided a-priori, e.g., hammer heads are made of wood or metal (Shown in **Table 1**). Note that this information can also be obtained using common knowledge bases such as RoboCSE (Daruna et al., 2019).

For material scoring, we seek to train models that can predict whether an input material is suited for performing a specific action. We represent the material properties of the object using spectral readings, since it has been shown to work well for material classification problems in prior work (Erickson et al., 2019, 2020; Shrivatsav et al., 2019). For extracting the spectral readings, the robot uses a commercially available handheld spectrometer[4], called SCiO, to measure the reflected intensities of different wavelengths, in order to profile and classify object materials. The spectrometer generates a 331-D real-valued vector of spectral intensities. Then given a dataset of SCiO measurements from an assortment of objects, we train a model through supervised learning to output a class label indicating the material of the object.

For the material score prediction, given the spectral readings for the action parts in O_a denoted by $\mathcal{K}$, we map the predicted class label to values in **Table 1** to compute the material score using the prediction confidence of the model. Let $T(a)$ denote the set of appropriate materials for performing an action a. Then the material score is computed as:

$$\phi^s_{mat}(O_a) = \begin{cases} z = \prod_{o_i \in \mathcal{K}} \max_{c_i \in T(a)} p(c_i|o_i), & \text{if } z \geq \text{t} \\ -\infty, & \text{otherwise} \end{cases} \tag{2}$$

Where, p is the prediction confidence of the network regarding the class c_i. We compute the max prediction confidence across all the appropriate classes $c_i \in T(a)$, and their product over the action parts in $\mathcal{K}$. For example, for the action "join-hammer(O_a)," where O_a consists of two objects (o_1, o_2), the material score $\phi^s_{mat}(O_a) = \max(p(metal|o_1), p(wood|o_1))$. If the max value exceeds some threshold[5] denoted by t, then the corresponding value is returned. Otherwise, the model returns $-\infty$. Hence, note that material prediction acts as a hard constraint, by directly eliminating any objects that are made of inappropriate materials, thus reducing the potential search effort.

3.3.3. Attachment Prediction

Given a set of objects, we seek to predict whether the objects can be attached to construct a tool. This is indicated by the *canAttach*() function in Algorithm 2. In order to attach the objects, we consider three attachment types for creating fixed attachments, namely, *pierce attachment* (piercing one object with another, e.g., foam pierced with a screwdriver), *grasp attachment* (grasping one object with another, e.g., a coin grasped with pliers), and *magnetic attachment* (attaching objects via magnets on them). For magnetic attachments, we manually specify whether magnets are present on the objects, enabling them to be attached. For pierce and grasp attachment, we check whether the attachments are possible as described below. If no attachments are possible for the given set of objects, the feature score returns $-\infty$, indicating that the objects are not a viable combination. Thus, the search eliminates any objects that cannot be attached.

3.3.3.1. Pierce attachment

Similar to material reasoning, we use the SCiO sensor to reason about material pierceability. We assume homogeneity of materials, i.e., if an object is pierceable, it is uniformly pierceable throughout the object. We train a neural network to output a binary label indicating pierceability of the input spectral reading (Nair et al., 2019b). If the model outputs 0, the objects cannot be attached via piercing.

3.3.3.2. Grasp attachment

To predict grasp attachment, we model the grasping tool (pliers or tongs) as an extended robot gripper. This allows the use of existing robot grasp sampling approaches (Zech and Piater, 2016; ten Pas et al., 2017; Levine et al., 2018), for computing locations where a given object can be grasped. In particular, we use the approach discussed by ten Pas et al., that outputs a set of grasp locations given the input parameters reflecting the attributes of the pliers or tongs used for grasping (ten Pas et al., 2017). If the approach could not sample any grasp locations, the objects cannot be attached via grasping.

3.4. Incorporating the Sensor Trust Parameter

In this section, we describe how the sensor trust parameter (Line 4, Algorithm 2) is incorporated to compute the feature score in two different ways. The first approach includes trusting the shape, material, and attachment predictions of the models described above. The second approach allows the robot to deal with possible false negatives in material and attachment predictions, by only incorporating the shape score into the feature score computation.

3.4.1. Fully Trust Sensors

In the case that the robot fully trusts the material and attachment predictions, the trust parameter is set to *true* (Line 4, Algorithm 2). The final feature score is then computed as a weighted sum of the shape and material scores, if the objects can be attached (Algorithm 2, Lines 5–8). We found uniform weights of $\lambda_1 = 1, \lambda_2 = 1$, to work well for tool constructions. If the objects cannot be attached, then $\phi(s) = -\infty$, indicating that the

[4] https://www.consumerphysics.com/ - Note that SCiO can be controlled via an app that enables easy scanning of objects. The robot simply moves the scanner over the object, and the user presses a key within the app to scan the object.
[5] We empirically determined a threshold of 0.6 to work well.

objects in O_a do not form a valid combination. Otherwise, using Equations (1) and (2):

$$score(s, O_a) = \lambda_1 * \phi^s_{shape}(O_a) + \lambda_2 * \phi^s_{mat}(O_a) \quad (3)$$

Since both material and attachment predictions are hard constraints, certain object combinations can be assigned a score of $-\infty$, indicating that the robot does not attempt these constructions. As described before, this can lead to cases of false negatives where the robot is unable to find the correct construction due to incorrect material or attachment predictions. In these cases, the algorithm tracks the rejected object combinations in O_{reject} (Algorithm 1, Line 26), and repeats the search as described below, by switching trust to *false* (Algorithm 1, Lines 33–34).

3.4.2. Not Fully Trust Sensors

In case of false negatives, the robot can choose to eliminate the hard constraints of material and attachment prediction from the feature score computation, thus allowing the robot to explore the initially rejected object combinations by using only the shape score. This is achieved by setting the trust flag to *false* in our implementation (Lines 14–15, Algorithm 2). In this case, we attempt to re-plan using the feature score as:

$$\phi(s) = \begin{cases} \phi^s_{shape}(O_a), & \text{if } O_a \subseteq O_{reject} \\ -\infty, & \text{otherwise} \end{cases} \quad (4)$$

Here, O_{reject} indicates the set of objects that were initially rejected by the material and/or the attachment predictions. Since, shape score is a soft constraint, i.e., it does not eliminate any object combinations completely, we use the shape score to guide the search in case of the rejected objects. In the worst case, this causes the robot to explore all ${}^nP_m$ permutations of objects. However, as shown in our results, shape score can serve as a useful guide for improving tool construction performance in practice, when compared to naively exploring all possible object combinations. The final feature score computation, influenced by attachments and the trust parameter, can be summarized as follows from Equations (3), (4):

$$\phi(s) = \begin{cases} score(s, O_a), & \text{if attachable \& trust} \\ \phi^s_{shape}(O_a), & \text{if } \textbf{not} \text{ trust \& } O_a \subseteq O_{reject}. \\ -\infty, & \text{otherwise} \end{cases}$$

3.5. Final Cost Computation

Once the feature score is computed, the final cost function is computed as $f(s) = g(s) + h(s) - \lambda * \phi(s)$. Interestingly, we found that $\lambda = 1$, thus $f(s) = g(s) + h(s) - \phi(s)$, performs very well with the choice of search algorithms and heuristics in this work for the problem of tool construction. In this case, the higher the feature score $\phi(s)$, the lower the cost $f(s)$, in turn guiding the search to choose nodes with higher feature score (lower $f(s)$ values). Additionally, the values of the feature score are within the range $0 \leq \phi(s) \leq 2$. Since we use existing planning heuristics that have been shown to work well, and the task plans generated have $\gg$2 steps involved, $g(s)+h(s) \gg 2$ and thus, $f(s) > 0$. Thus, $\lambda = 1$ works well for the problems described in this work. However, this presents an interesting research question for our future work in terms of an in-depth analysis of the choice of heuristic and feature score values, and its influence on the guarantees of the search.

3.6. Implementation Details

In this section, we describe additional details regarding the implementation of the work, both in terms of the algorithm, as well as the physical implementation on the robot.

3.6.1. Algorithm Implementation

In terms of implementation, the process begins with the trust parameter set to *True*. FGS generates a task plan that involves combining objects to construct the required tool. Once a task plan is successfully found, the robot can proceed with executing the task plan and joining the parts indicated by O_a as described in Nair et al. (2019b), to construct the required tool for completing the task. If the tool could not be successfully constructed or used, the plan execution is said to have failed, and the robot re-plans to generate a new task plan with a different object combination, since the algorithm tracks the attempted object combinations. Note that the approach also keeps track of object combinations rejected by material and attachment predictions in O_{reject}. If no solution could be found with trust set to *True*, and $O_{reject} \neq \emptyset$, then the robot switches trust to *false*, and FGS explores the object combinations within O_{reject} (Lines 33-34 of Algorithm 1). If no solution could be found with either trust setting, FGS returns a complete failure and terminates.

Further note that in this work, we do not explicitly deal with symbol grounding (Harnad, 1990) and symbol anchoring (Coradeschi and Saffiotti, 2003) problems. We overcome these issues by manually mapping the object point clouds to their specific symbols within the planning domain definition. Once the task plan is generated, the mapping is then used to match the symbols within the task plan to their corresponding objects in the physical world, via their point clouds. However, existing approaches can potentially be adapted in order to refine the symbol grounding functions (Hiraoka et al., 2018), or to enable the robot to automatically extract the relationships between the object point clouds and their abstract symbolic representations (Konidaris et al., 2018).

3.6.2. Physical Implementation

The spectrometer used in this work can be activated either using a physical button located on the sensor, or through an app that is provided with the sensor. However, pressing the physical button requires precision and careful application of the correct amount of force, which can be challenging for the robot since it may potentially damage the sensor if the applied force exceeds a certain threshold. To prevent this, in our implementation, the robot simply moves the scanner over the objects, and the user then manually presses a key within the app to activate the sensor. Additionally, the rate of scanning is also limited by the speed of the robot arm itself. Since the robot arm used in this work moves rather slowly, it took about ≈1.7 min on average to scan 10 objects, while this would take <30 s for a human. Overcoming these issues and several other manipulation challenges are essential to ensure practical applicability of this work. We discuss this in more detail in section 5.

4. EXPERIMENTAL VALIDATION AND RESULTS

In this section, we describe our experimental setup and present our results alongside each evaluation. We validate our approach on three diverse types of tasks involving tool construction in a household domain, namely, wood-working, cooking, and cleaning. For wood-working tasks, the tools to be constructed include hammer and screwdriver; for cooking tasks the tools include spatula and ladle; and lastly for cleaning tasks the tools include rake and squeegee. Each tool is constructed from two parts ($m = 2$) corresponding to the action and grasp parts of the tool[6]. Our experiments seek to validate the following three aspects:

1. **Performance of feature guided A^* against baselines:** In order to investigate the informativeness of including feature score in heuristic search, we evaluate the feature guided A^* approach against three baselines. We also evaluate our approach in terms of the two different settings of the sensor trust parameter to investigate the benefits of introducing sensor trust.
2. **Combining feature scoring with other heuristic search algorithms:** To investigate whether feature scoring can generalize to other search approaches, we integrate feature scoring with two additional heuristic search algorithms. Specifically, we present results combining feature scoring with weighted A^* and enforced hill-climbing with the fast-forward heuristic (Hoffmann and Nebel, 2001).
3. **Adaptability of task plans to objects in the robot's environment:** We evaluate whether the robot can adapt its task plans to appropriately use the constructed tool, as the objects available to the robot for tool construction are modified. This measures whether the robot can flexibly generate task plans in response to the objects in the environment.

For all our experiments, we use a test set consisting of 58 previously unseen candidate objects for tool construction (shown in **Figure 1**). These objects consist of metal (11/58), wood (12/58), plastic (19/58), paper (2/58), and foam (14/58) objects. Only the foam and paper objects are pierceable. Prior to planning, the robot scans the materials of the objects for material scoring and attachment predictions. For our results, we evaluate the statistical significance where it is applicable, using repeated measures ANOVA with *post-hoc* Tukey's test. We discuss each experiment in more detail below, along with the results for each.

4.1. Performance of Feature Guided A*

In this section, we evaluate the performance of feature guided A^* against three baselines: (i) standard A^*, where $f(s) = g(s) + h(s)$, (ii) feature guided uniform cost search, where $f(s) = g(s) + 2.0 - \phi(s)$, and (iii) standard uniform cost search, where $f(s) = g(s)$. In (ii), we use $2.0 - \phi(s)$ to add a positive value to $g(s)$ since, $0 \leq \phi(s) \leq 2$. As a heuristic with A^*, we use the cost optimal landmark heuristic (Karpas and Domshlak, 2009). We also vary the sensor trust parameter, and present results for the two cases where the robot is not allowed to change the trust parameter (trust always set to *true*, i.e., lines 33–34 of Algorithm 1 not executed), and for the case where the robot is allowed to change it to *false* when no plan is found.

For the evaluation, we create six different tasks, two tasks each for wood-working, cooking and cleaning. Each task requires the construction of *one* specific tool for its completion, e.g., one of the tasks in wood-working requires construction of a hammer, and the other requires construction of a screwdriver. For each task we created 10 test cases, where each test case consisted of 10 objects chosen from the 58 in **Figure 1**, that could potentially be combined to construct the required tool. We report the average results across the test cases for each task type (total 10×2 cases per task type with 10 candidate objects per case). We create each test set by choosing a random set of objects, ensuring that only one "correct" combination of objects exists per set. The correct combinations are determined based on human assessment of the objects. For each task, we instantiate the corresponding domain and problem definitions in PDDL[7].

The metrics used in this experiment include (i) the *number of nodes expanded* during search as a measure of computational resources consumed, (ii) the *number of failed construction attempts* before a working tool was found (also referred to as "attempts" in this paper), and (iii) the *success rate* indicating the number of times the robot successfully found a working tool. Ideally, we would like the number of nodes expanded and the number of failed construction attempts to be as low as possible. Note that the brute force number of failed construction attempts for 10 objects is 89, since there are ${}^{10}P_2$ possible object permutations for $m = 2$, with 89 incorrect possibilities. Ideally, we would like the number of failed construction attempts to be 0. The success rate should be as high as possible, ideally equal to 100%.

Table 2 shows the performance of feature guided A^* (where $f(s) = g(s) + h(s) - \phi(s)$, denoted by "FS+H") compared to the different baselines: "H" denotes standard A^* (where $f(s) = g(s) + h(s)$), "FS" denotes feature guided uniform cost search (where $f(s) = g(s) + 2.0 - \phi(s)$), and "UCS" denotes standard uniform cost search (where $f(s) = g(s)$). The values reported per task are the average performances across the test cases where tool constructions were successful. As shown in **Table 2**, incorporating feature scoring (FS, FS+H) helps significantly reduce the number of failed construction attempts compared to the baselines without feature scoring (H, UCS), with $p < 0.001$. Since heuristics can help reduce the search effort in terms of number of nodes expanded, we see that approaches that do not use heuristics (FS and UCS) expand significantly more

[6] In this work, we pre-specify the trajectories to be followed when combining the objects to construct the tool.

[7] In the planning problem definition, the objects are instantiated numerically through "obj0" to "obj9," where each literal is manually assigned to one of the 10 objects during planning time. Our planning and domain definitions are available at: https://github.com/Lnair1993/Tool_Macgyvering.

FIGURE 1 | Dataset of 58 objects used for the experiments, made of different materials.

TABLE 2 | Table comparing feature guided A^* ("FS+H") with baselines.

	Cleaning				Cooking				Wood-working			
	FS+H	**H**	**FS**	**UCS**	**FS+H**	**H**	**FS**	**UCS**	**FS+H**	**H**	**FS**	**UCS**
# Nodes	**5187**	**5187**	9061	9061	**329**	604	36237	36213	7264	**6936**	28606	28734
# Failed attempts	**2**	46	3	49	**3**	48	4	40	**2**	45	**2**	37

The other notations: "H"—standard A; "FS"— feature guided uniform cost search; "UCS"—standard uniform cost search. This table reports the average number of failed attempts per task, across test cases where tool construction was successful. Note that the max number of failed attempts possible is 89 (brute force). Bold highlights the best performance values for the different metrics.*

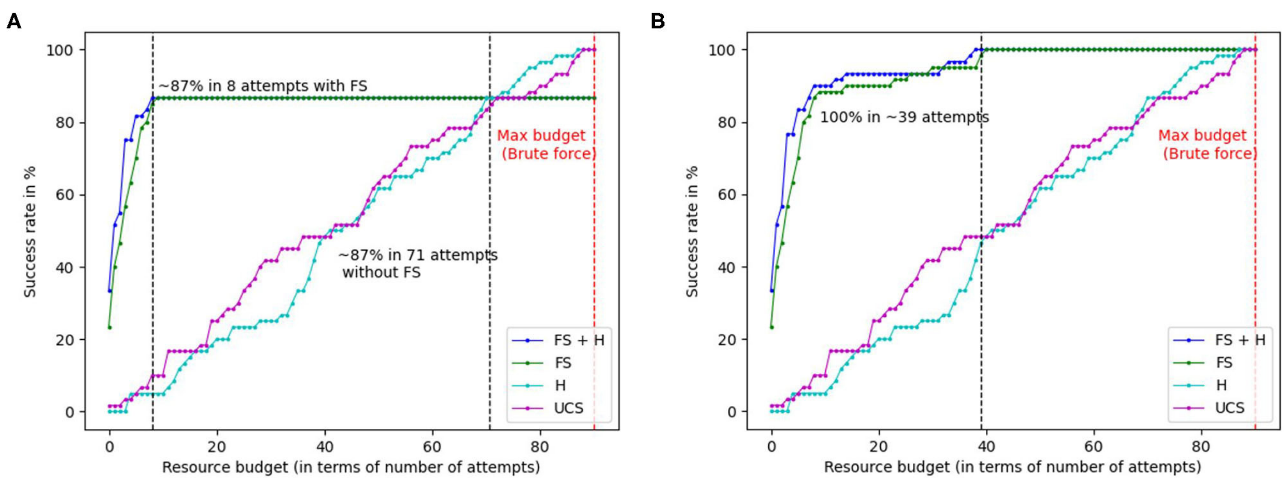


FIGURE 2 | Graphs highlighting the success rates for the two different modes of feature scoring based on sensor trust parameter, in relation to the number of failed attempts. Note that X-axis highlights the *actual* number of attempts across all test cases for wood-working, cooking, and cleaning put together. **(A)** Graph showing the success rate compared to the number of attempts when sensors are fully trusted. **(B)** Graph showing the success rate vs. the number of attempts when sensors are not fully trusted.

nodes than FS+H and H, with $p < 0.001$. Note that there is no statistically significant difference in the number of nodes expanded between H and FS+H. Thus, using feature scoring with heuristics (FS+H) yields the best performance in terms of *both* number of nodes expanded, and the number of failed construction attempts. To summarize, these results show that feature scoring is informative to heuristic search by significantly reducing the average number of failed construction attempts to ≈ 2 compared to ≈ 46 without it (brute force number of failed attempts is 89).

Further, in **Figures 2A,B**, we plot the success rate vs. the resource budget of the robot in terms of the *permissible* number of failed attempts. That is, the robot is not allowed to try any more than a fixed number of attempts, indicated by the resource budget. **Figure 2A** considers the case where the sensor trust parameter is always set to *true*, and **Figure 2B** considers the case when the robot is allowed to switch the trust to *false*, if a solution was not found. Note that in contrast to **Table 2**, the graphs report *actual* number of failed attempts, *across all tasks*, whereas **Table 2** reports the *average* number of failed attempts across the test cases per task, for tool constructions that were successful. In **Figure 2A**, we see that FGS (FS+H and FS) achieves a success rate of 86.67% (52/60 constructions) within a resource budget of ≈ 8 failed attempts to do so. This indicates that 13.33% of the valid constructions were treated as false negatives by material and attachment predictions, and were completely removed from consideration (unattempted). Thus, increasing the permissible resource budget beyond 8, does not make any difference. Without feature scoring, H and UCS achieve a success rate of 87% with a budget of 71 attempts, and 100% after exploring nearly every possible construction (max resource budget of 89 failed attempts). In contrast, when the robot is allowed to switch the trust parameter, the robot uses shape scoring alone to continue guiding the search. As shown in **Figure 2B**, FGS (FS+H and FS) achieves 100% success rate within a budget of ≈ 39 attempts, since the robot does not eliminate any object combinations from consideration. The performance is also significantly better than the baselines that do not use feature scoring. This is because shape scoring guides the search through the space of object combinations based on the objects' shape fitness, compared to H and UCS that do not have any measure of the fitness of the objects for tool construction. To summarize, feature scoring enables the robot to successfully construct tools by leveraging the sensor trust parameter, while significantly outperforming the baselines in terms of the resource budget required.

In order to understand which tools were more challenging for feature scoring, **Table 3** shows a tool-wise breakdown in performance for feature guided A^* for the two different sensor trust values. The notation "trust" denotes the case where sensors are fully trusted, and "$\sim$trust" denotes case where they are not fully trusted. When the sensors are fully trusted, rakes were a particularly challenging test case, as indicated by the lowest success rate of 7/10. In contrast, hammers and ladles have a success rate of 10/10. The failure cases for each tool arises from incorrect material and attachment predictions. While not fully trusting the sensors ($\sim$trust) leads to a 100% success rate (60/60 cases), using shape score alone leads to more failed construction attempts when compared to combining shape with material and attachment predictions since shape alone is less informative (e.g., for rake, $\sim$trust has 7 failed attempts vs. 3 failed attempts for trust).

Figure 3 shows sample task plans generated by the robot in cooking and cleaning tasks. In the case of cooking, the robot needed a spatula to flip the eggs, and used a flat piece (obj4) with tongs (obj5) to construct the spatula via grasp attachment. For cleaning, the robot needed a squeegee to clean the window, and used a foam block (obj1) and screwdriver (obj6) to construct the squeegee via pierce attachment. Without the constructed tools, the actions highlighted in red would fail, i.e., the "flip" action would fail without the constructed spatula. Hence, FGS enables the robot to replace missing tools through construction. To summarize, the key findings of this experiment indicate that feature scoring is highly informative for heuristic search by reducing the number of nodes expanded by $\approx 82\%$, and the number of failed construction attempts by $\approx 93\%$, compared to the baselines. Further, allowing the robot to switch the trust parameter when a plan is not found, helps achieve a success rate of 100% within a budget of ≈ 39 attempts, significantly outperforming baselines that do not use feature scoring.

4.2. Feature Scoring With Other Heuristic Search Algorithms

To demonstrate that feature scoring generalizes to other search approaches, in this section we present results for combining feature scoring with weighted A^* (Pohl, 1970), and enforced hill-climbing using the fast-forward heuristic (Hoffmann and Nebel, 2001). We use the same experimental setup and metrics as described in section 4.1. In addition, we also measure the output plan length to investigate the optimality of the different approaches. For weighted A^*, feature scoring is incorporated as $f(s) = g(s) + w * (h(s) - \phi(s))$, where w indicates a weight parameter[8]. For enforced hill-climbing, the cost function is computed as $f(s) = h(s) - \phi(s)$. For both weighted A^* and enforced hill-climbing, we use the fast-forward heuristic, which has been shown to be successful for planning tasks in prior work (Hoffmann and Nebel, 2001).

In **Table 4**, we present the results for feature scoring combined with A^* and the cost-optimal landmark heuristic ("A^*+LM"), weighted A^* with fast-forward heuristic ("wA^*+FF"), and enforced hill-climbing with fast forward heuristic ("eHC+FF"). Compared to A^*+LM, wA^*+FF and eHC+FF reduce the computational effort (fewer nodes expanded) in return for sub-optimal solutions (longer plan lengths). This is expected of weighted A^* and enforced hill-climbing since they are inadmissible algorithms. There is no statistically significant difference between # failed construction attempts in each case. To summarize, the key finding of this experiment is that feature scoring can be applied to other planning heuristics such as fast-forward, and other heuristic search algorithms like weighted A^* and enforced hill-climbing, to further reduce computational effort, albeit at the cost of optimality in terms of plan length.

[8]Weight was set to 5.0.

TABLE 3 | Table showing tool-wise breakdown in performance for feature guided A^*.

	Cleaning				Wood-working				Cooking			
	Squeegee		Rake		Hammer		Screwdriver		Spatula		Ladle	
	Trust	∼Trust	Trust	∼Trust	Trust	∼Trust	Trust	∼Trust	Trust	∼Trust	Trust	∼Trust
# Failed attempts	0	1	3	7	2	2	2	8	3	7	2	2
# Success	9/10	10/10	7/10	10/10	10/10	10/10	8/10	10/10	8/10	10/10	10/10	10/10

This table reports the average number of failed attempts per tool, across cases where tool construction was successful. The notation ∼trust indicates cases where sensors are not fully trusted. Note that max # failed attempts is 89.

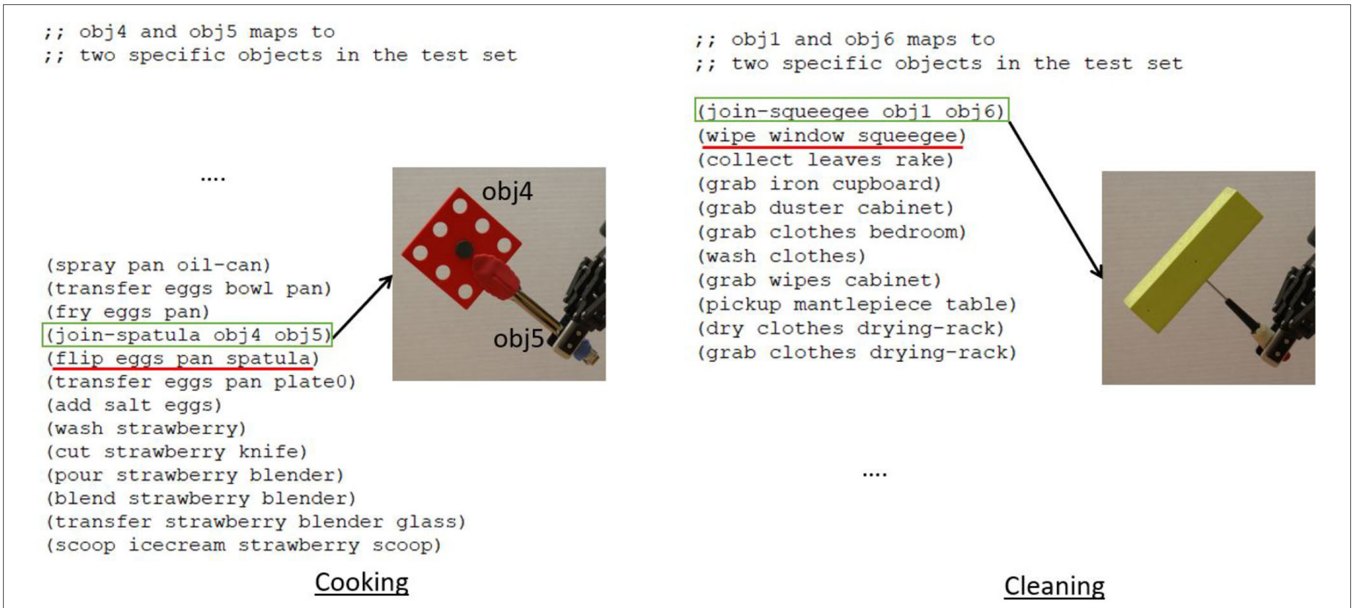


FIGURE 3 | (Left) A sample task plan where a spatula must be constructed for a cooking task, and the planner uses the flat piece (obj4 in the problem definition), and tongs (obj5 in the problem definition). The action "join-spatula" refers to the construction of the spatula using obj4 and obj5. Similarly, **(right)** a squeegee is constructed from obj1 (foam block) and obj6 (screwdriver) for the cleaning task. Without tool construction (highlighted in green) the actions underlined in red would fail.

TABLE 4 | Table showing performance of feature guided Weighted A^* (wA^*) and feature guided Enforced Hill-Climbing (eHC) with the fast-forward heuristic (FF).

	Cleaning			Cooking			Wood-working		
	A^*+LM	wA^*+FF	eHC+FF	A^*+LM	wA^*+FF	eHC+FF	A^*+LM	wA^*+FF	eHC+FF
# Nodes	5187	21	21	329	23	35	7264	25	38
# Failed attempts	2	2	4	3	3	4	2	1	2
Plan length	20	22	22	19	19	19	11	15	15

4.3. Adaptability of Task Plans

In this section we evaluate the adaptability of our FGS approach to generate task plans based on objects in the environment, to appropriately use the constructed tool. We create three tasks, one task each for wood-working, cooking, and cleaning. In each of the tasks, *either* of two tools can be constructed to successfully complete the task, but there is only one ground truth depending on the objects available for construction. That is, the available objects only enable the construction of one of the two tools. Thus, the robot has to *correctly choose the tool to be constructed*. In addition, the robot must adapt the task plan to appropriately use the constructed tool. For the wood-working task either a hammer (with action "hit") or a screwdriver (with action "tighten") can be used to attach two pieces of wood; for the cooking task either a spatula (with action "flip") or a ladle (with action "scoop") can be used to flip eggs; and for the cleaning task, either a squeegee (with action "reach") or a rake (with action "collect") can be used to collect garbage.

For the evaluation, we create three different tasks, one each in wood-working, cooking, and cleaning. For each task, either

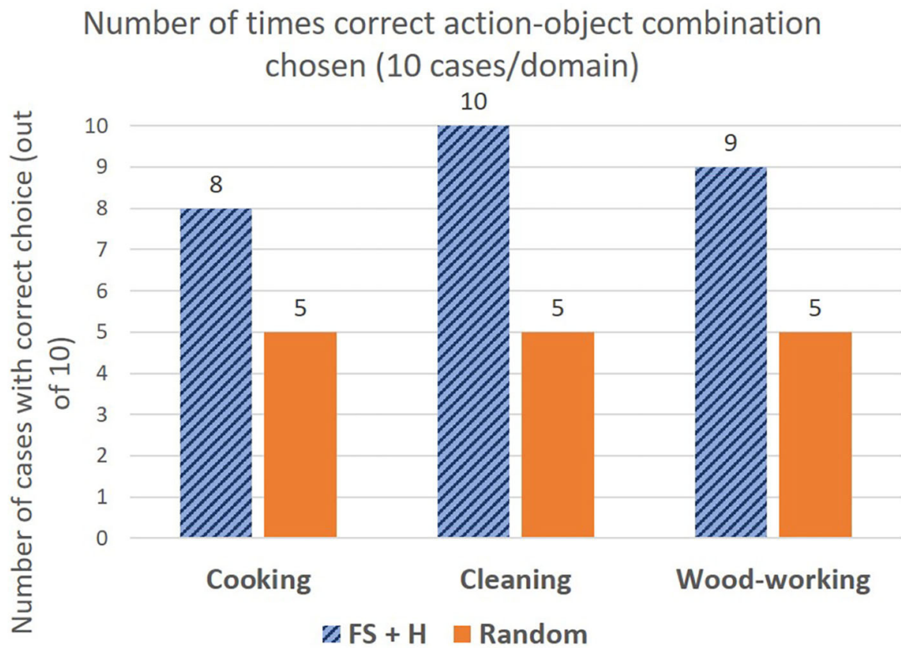


FIGURE 4 | Graph highlighting the number of times the correct object combination was chosen, compared to the random selection baseline. FGS significantly outperforms random baseline ($p < 0.01$).

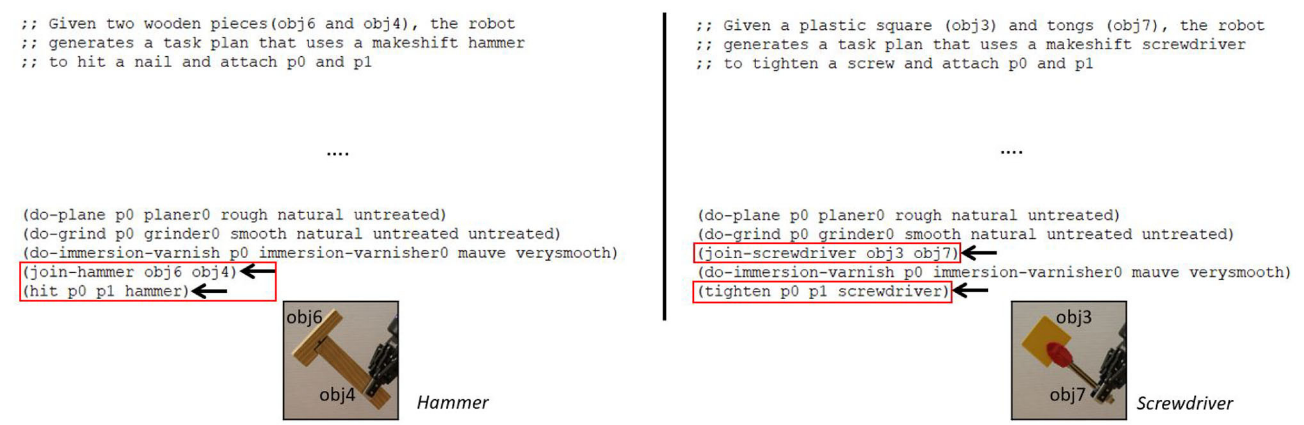


FIGURE 5 | This figure shows the results for two of the test cases in wood-working. The task plans are adapted based on the constructed tool (i.e., hammer or screwdriver), to either "hit" or "tighten" to attach the two pieces of wood $p0$ and $p1$. Arrows denote the parts of the task plan that are adapted.

one of two tools can be used to complete the task as described above. For each task, we created 10 different test sets of random objects, similar to the experiment described in section 4.1. In each case, only one "correct" combination exists. Thus, the robot has to correctly identify which of the two tools can be constructed for accomplishing the task, given the set of objects. We evaluate the performance of feature guided A^* in each case alongside a random selection baseline to demonstrate the difficulty of the problem. The random selection baseline randomly chooses one of the two tool construction options for each task. Note that for each task, the domain and problem definitions are unchanged across the 10 test cases of objects. This indicates that the task plan adaptability does not require any manual modifications by the user, instead is the direct result of the sensor inputs received by the robot.

The key metric used in this experiment includes the number of times the robot chose the correct tool to construct for each task. Thus, if the robot chose to construct a hammer, when the correct combination was to construct a screwdriver, the attempt is considered to have failed. We also present qualitative results showing some of the sample task plans and tools constructed by the robot for different sets of objects.

FIGURE 6 | Collage indicating sample tool constructions output for two test cases per task. The solid and dashed brackets indicate the test set of objects provided in each case, along with the tool constructed for it. As the objects are changed, the corresponding constructed tool and action is different. Note that the problem and domain definitions are fixed for each task, and unchanged across the test cases per task.

Figure 4 shows the performance of feature guided A^* compared to the random selection baseline. We see that feature guided A^* chooses the correct tool for 27/30 cases, and significantly outperforms the random selection baseline ($p < 0.01$). The failure cases in the wood-working task arise due to noisy material detection. In the case of cooking task, the noisy point clouds sensed by the RGBD camera leads to incorrect choices, e.g., the concavity of bowls was not correctly detected for some ladles.

In **Figure 5**, we show two task plans that are generated within the task of wood-working. For the same task, either a hammer or a screwdriver can be used to attach two pieces of wood $p0$ and $p1$. Depending, on the objects available in the environment, the robot chooses to construct one of the two tools and adapts the task plan to use the corresponding tool for completing the task. As shown in the left of **Figure 5**, the robot chose to construct a hammer to "hit" and attach the two pieces of wood. Whereas, shown in the right of **Figure 5**, the robot chose to construct a screwdriver to "tighten" and attach the two pieces of wood. Similar adaptations are observed for the remaining two tasks as well: "scoop" with ladles vs. "flip" with spatulas in the cooking task, and "reach" with squeegees vs. "collect" with rakes in the cleaning task. Thus, the constructed tool depends on the objects in the environment, which in turn adapts the generated task plan to appropriately use the constructed replacement tool.

In **Figure 6**, we present some qualitative results for six different tools constructed by the robot for six of the test cases. The solid and dashed parentheses highlight the input test set. For example, given the metal bowl and metal pliers, the robot chooses to construct a ladle (and use the "scoop" action in the task plan). In contrast, when the pliers and bowl are replaced with a plastic handle and a flat plastic piece, the robot chooses to construct a spatula instead (and use the "flip" action in the task plan). Given that the problem and domain definitions are unchanged for the two cases, this shows that the robot is able to adapt the task plan in response to the objects in the environment. To summarize, the key finding of this experiment is that the robot is able to successfully adapt the task plan to construct and use the appropriate tool depending on the objects available for construction, with an accuracy of 90% (27/30 cases).

5. CONCLUSION AND FUTURE WORK

In this work, we presented the Feature Guided Search (FGS) approach that allows existing heuristic search algorithms to be efficiently applied to the problem of tool construction in the context of task planning. Our approach enables the robot to effectively construct and use tools in cases where the required tools for performing the task are unavailable. We relaxed key assumptions of the prior work in terms of eliminating the need to specify an input action, instead integrating tool construction within a task planning framework. Our key findings can be summarized as follows:

- FGS significantly reduces the number of nodes expanded by $\approx$ 82%, and the number of construction attempts by $\approx$ 93%, compared to standard heuristic search baselines.
- The approach achieves a success rate of 87% within a resource budget of 8 attempts when sensors are fully trusted, and 100% within a budget of 39 attempts, when the sensors are not fully trusted.
- FGS enables flexible generation of task plans based on objects in the environment, by adapting the task plan to appropriately use the constructed tool.
- Feature scoring can also be effectively combined with other heuristic search algorithms such as weighted A^* and enforced hill-climbing.

Our work is one of the first to integrate tool construction within a task planning framework, but there remain many unaddressed manipulation challenges in tool construction that are beyond the scope of this paper. Tool construction is a challenging manipulation problem that involves appropriately grasping and combining the objects to successfully construct the tool. That is, once the robot has correctly identified the objects that need to be combined (focus of this paper), the robot would then have to physically combine the objects, and use the constructed tool for the task. Currently, our work pre-specifies the trajectories to be followed for tool construction, although existing research in robot assembly can be leveraged to potentially accomplish this (Thomas et al., 2018). Further, a key question to be addressed is, *how can the robot learn to appropriately use the constructed tool?* Future work could address this problem by leveraging existing research in tool use (Stoytchev, 2005; Sinapov and Stoytchev, 2007, 2008), and trajectory-based skill adaptation (Fitzgerald et al., 2014; Gajewski et al., 2019). Upon successful construction of the tool, the research problem reduces to that of using the tool appropriately. In this case, the robot can either learn how to use the tool as described in Stoytchev (2005), Sinapov and Stoytchev (2008, 2007) or, the robot can adapt previously known tool manipulation skills to the newly constructed tool as described in Fitzgerald et al. (2014) and Gajewski et al. (2019). Addressing these challenges is important to further ensure practical applicability of tool construction.

Additionally, creation of tools through the attachment types discussed in this work is currently restricted to a limited number of use cases, in which two objects that have the specific attachment capabilities already exist, and are available to the robot. In the future, we seek to expand to more diverse types of attachments, including gluing or welding the objects together, as well as creation of tools from deformable materials, in order to improve the usability of our work. We further seek to expand on this work by investigating the application of feature scoring to domains other than tool construction. In particular, we seek to investigate the different ways in which feature score can be effectively combined with the cost function for other domains involving tool-use such as tool substitution. While our proposed cost function is dependent on the values of the feature score and is shown to perform well for tool construction, it is important to further investigate the cost function and its influence on the guarantees of the search to allow for a more generalized application of FGS.

FGS enables the robot to perform high-level decision making with respect to the objects that must be combined in order to construct a required tool. In this work, we use physical sensors (RGBD sensors and SCiO spectrometer) that produce partial point clouds and noisy spectral scans, leading to some challenges that commonly arise in the real world. Nevertheless, there are several open research questions that need to be addressed before this work can be deployed in a real setting. Thus, FGS is the first step within a larger pipeline, and we envision this work to be complementary to existing frameworks that are aimed at resilient and creative task execution, such as Antunes et al. (2016) and Stückler et al. (2016). In summary, FGS presents a promising direction for dealing with tool-based problems in the area of creative problem solving.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://github.com/Lnair1993/Tool_Macgyvering.

AUTHOR CONTRIBUTIONS

LN and SC conceived of the presented idea. LN developed the theory, performed the computations, and carried out the experiment. LN wrote the manuscript with support from SC. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by NSF IIS 1564080 and ONR N000141612835.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Christopher G. Atkeson and Dr. Kalesha Bullard for their valuable feedback and insights on this work. Preprint of this paper is available (Nair and Chernova, 2020).

REFERENCES

Abelha, P., and Guerin, F. (2017). "Learning how a tool affords by simulating 3D models from the web," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Vancouver, BC), 4923–4929.

Abelha, P., Guerin, F., and Schoeler, M. (2016). "A model-based approach to finding substitute tools in 3d vision data," in *2016 IEEE International Conference on Robotics and Automation (ICRA)* (Stockholm), 2471–2478.

Agostini, A., Aein, M. J., Szedmak, S., Aksoy, E. E., Piater, J., and Würgütter, F. (2015). "Using structural bootstrapping for object substitution in robotic executions of human-like manipulation tasks," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Hamburg), 6479–6486.

Antunes, A., Jamone, L., Saponaro, G., Bernardino, A., and Ventura, R. (2016). "From human instructions to robot actions: formulation of goals, affordances and probabilistic planning," in *2016 IEEE International Conference on Robotics and Automation (ICRA)* (Stockholm), 5449–5454.

Atkeson, C. G., Benzun, P. B., Banerjee, N., Berenson, D., Bove, C. P., Cui, X., et al. (2018). "What happened at the Darpa robotics challenge finals," in *The DARPA Robotics Challenge Finals: Humanoid Robots to the Rescue* (Pittsburgh, PA: Springer), 667–684.

Bagchi, A., and Mahanti, A. (1983). Search algorithms under different kinds of heuristics-a comparative study. *J. ACM* 30, 1–21.

Beck, S. R., Apperly, I. A., Chappell, J., Guthrie, C., and Cutting, N. (2011). Making tools isn't child's play. *Cognition* 119, 301–306. doi: 10.1016/j.cognition.2011.01.003

Boteanu, A., Kent, D., Mohseni-Kabir, A., Rich, C., and Chernova, S. (2015). "Towards robot adaptability in new situations," in *2015 AAAI Fall Symposium Series* (Arlington, VA: AAAI Press).

Cass, S. (2005). Apollo 13, we have a solution. *IEEE Spectr. On-line* 4:1. Available online at: https://spectrum.ieee.org/tech-history/space-age/apollo-13-we-have-a-solution

Choi, D., Langley, P., and To, S. T. (2018). "Creating and using tools in a hybrid cognitive architecture," in *2018 AAAI Spring Symposium Series* (Arlington, VA).

Coradeschi, S., and Saffiotti, A. (2003). An introduction to the anchoring problem. *Robot. Auton. Syst*. 43, 85–96. doi: 10.1016/S0921-8890(03)00021-6

Daruna, A., Liu, W., Kira, Z., and Chetnova, S. (2019). "Robocse: Robot common sense embedding," in *2019 International Conference on Robotics and Automation (ICRA)* (Montreal, QC), 9777–9783.

Erdogan, C., and Stilman, M. (2013). "Planning in constraint space: automated design of functional structures," in *2013 IEEE International Conference on Robotics and Automation (ICRA)* (Karlsruhe), 1807–1812.

Erickson, Z., Luskey, N., Chernova, S., and Kemp, C. (2019). Classification of household materials via spectroscopy. *IEEE Robot. Autom. Lett*. 4, 700–707. doi: 10.1109/LRA.2019.2892593

Erickson, Z., Xing, E., Srirangam, B., Chernova, S., and Kemp, C. C. (2020). Multimodal material classification for robots using spectroscopy and high resolution texture imaging. *arXiv [Preprint]*. arXiv:2004.01160v2. Available online at: https://arxiv.org/abs/2004.01160

Fitzgerald, T., Goel, A. K., and Thomaz, A. L. (2014). "Representing skill demonstrations for adaptation and transfer," in *AAAI Symposium on Knowledge, Skill, and Behavior Transfer in Autonomous Robots* (Pao Alto, CA).

Freedman, R., Friedman, S., Musliner, D., and Pelican, M. (2020). "Creative problem solving through automated planning and analogy," in *AAAI 2020 Workshop on Generalization in Planning (GenPlan 20)* (New York, NY).

Gajewski, P., Ferreira, P., Bartels, G., Wang, C., Guerin, F., Indurkhya, B., et al. (2019). "Adapting everyday manipulation skills to varied scenarios," in *2019 International Conference on Robotics and Automation (ICRA)* (Montreal, QC), 1345–1351.

Gizzi, E., Castro, M. G., and Sinapov, J. (2019). "Creative problem solving by robots using action primitive discovery," in *2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)* (Oslo), 228–233.

Harnad, S. (1990). The symbol grounding problem. *Phys. D* 42, 335–346.

Hart, P. E., Nilsson, N. J., and Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Trans. Syst. Sci. Cybern*. 4, 100–107.

Hiraoka, T., Onishi, T., Imagawa, T., and Tsuruoka, Y. (2018). Refining manually-designed symbol grounding and high-level planning by policy gradients. *arXiv [Preprint]*. arXiv:1810.00177. Available online at: https://arxiv.org/abs/1810.00177

Hoffmann, J., and Nebel, B. (2001). The FF planning system: fast plan generation through heuristic search. *J. Artif. Intell. Res*. 14, 253–302. doi: 10.1613/jair.855

Jones, T. B., and Kamil, A. C. (1973). Tool-making and tool-using in the northern blue jay. *Science* 180, 1076–1078.

Karpas, E., and Domshlak, C. (2009). "Cost-optimal planning with landmarks," in *IJCAI* (Pasadena, CA), 1728–1733.

Konidaris, G., Kaelbling, L. P., and Lozano-Perez, T. (2018). From skills to symbols: learning symbolic representations for abstract high-level planning. *J. Artif. Intell. Res*. 61, 215–289. doi: 10.1613/jair.5575

Levihn, M., and Stilman, M. (2014). "Using environment objects as tools: unconventional door opening," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014)* (Chicago, IL), 2502–2508.

Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., and Quillen, D. (2018). Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *Int. J. Robot. Res*. 37, 421–436. doi: 10.1177/0278364917710318

McDermott, D., Ghallab, M., Howe, A., Knoblock, C., Ram, A., Veloso, M., et al. (1998). *PDDL- The Planning Domain Definition Language*. Technical Report CVC TR-98-003/DCS TR-1165, Yale Center for Computational Vision and Control.

Myers, A., Teo, C. L., Fermüller, C., and Aloimonos, Y. (2015). "Affordance detection of tool parts from geometric features," in *International Conference on Robotics and Automation (ICRA)* (Seattle, WA), 1374–1381.

Nair, L., Balloch, J., and Chernova, S. (2019a). "Tool macgyvering: tool construction using geometric reasoning," in *International Conference on Robotics and Automation (ICRA)* (Montreal, QC), 5837–5843.

Nair, L., and Chernova, S. (2020). Feature guided search for creative problem solving through tool construction. *arXiv [Preprint]*. arXiv:2008.10685. Available online at: https://arxiv.org/abs/2008.10685

Nair, L., Srikanth, N., Erikson, Z., and Chernova, S. (2019b). "Autonomous tool construction using part shape and attachment prediction," in *Proceedings of Robotics: Science and Systems* (Messe Frieburg), 1–10.

Nyga, D., Roy, S., Paul, R., Park, D., Pomarlan, M., Beetz, M., et al. (2018). "Grounding robot plans from natural language instructions with incomplete world knowledge," in *Conference on Robot Learning* (Zurich), 714–723.

Olteteanu, A.-M., and Falomir, Z. (2016). Object replacement and object composition in a creative cognitive system. Towards a computational solver of the alternative uses test. *Cogn. Syst. Res*. 39, 15–32. doi: 10.1016/j.cogsys.2015.12.011

Pierson, A., and Schwager, M. (2016). "Adaptive inter-robot trust for robust multi-robot sensor coverage," in *Robotics Research* (Springer), 167–183. Available online at: https://link.springer.com/chapter/10.1007/978-3-319-28872-7_10

Pohl, I. (1970). Heuristic search viewed as path finding in a graph. *Artif. Intell*. 1, 193–204.

Saboia, M., Thangavelu, V., Gosrich, W., and Napp, N. (2018). Autonomous adaptive modification of unstructured environments. *Robot. Sci. Syst*. 14, 70–78. doi: 10.15607/RSS.2018.XIV.070

Sarathy, V. (2018). Real world problem-solving. *Front. Hum. Neurosci*. 12:261. doi: 10.3389/fnhum.2018.00261

Sarathy, V., and Scheutz, M. (2017). The MacGyver test-a framework for evaluating machine resourcefulness and creative problem solving. *arXiv [Preprint]*. arXiv:1704.08350. Available online at: https://arxiv.org/abs/1704.08350

Sarathy, V., and Scheutz, M. (2018). MacGyver problems: AI challenges for testing resourcefulness and creativity. *Adv. Cogn. Syst.* 6.

Schoeler, M., and Wörgötter, F. (2015). Bootstrapping the semantics of tools: affordance analysis of real world objects on a per-part basis. *IEEE Trans. Cogn. Dev. Syst*. 8, 84–98. doi: 10.1109/TAMD.2015.2488284

Shrivatsav, N., Nair, L., and Chernova, S. (2019). Tool substitution with shape and material reasoning using dual neural networks. *arXiv [Preprint]*. arXiv:1911.04521.

Sinapov, J., and Stoytchev, A. (2007). "Learning and generalization of behavior-grounded tool affordances," in *2007 IEEE 6th International Conference on Development and Learning* (London), 19–24.

Sinapov, J., and Stoytchev, A. (2008). "Detecting the functional similarities between tools using a hierarchical representation of outcomes," in *2008 7th IEEE International Conference on Development and Learning* (Monterey, CA), 91–96.

Stilman, M., Zafar, M., Erdogan, C., Hou, P., Reynolds-Haertle, S., and Tracy, G. (2014). Robots using environment objects as tools the 'MacGyver' paradigm for mobile manipulation," in *2014 IEEE International Conference on Robotics and Automation (ICRA)* (Hong Kong), 2568–2568.

Stout, D. (2011). Stone toolmaking and the evolution of human culture and cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. 366, 1050–1059. doi: 10.1098/rstb.2010.0369

Stoytchev, A. (2005). "Behavior-grounded representation of tool affordances," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation* (Barcelona), 3060–3065.

Stückler, J., Schwarz, M., and Behnke, S. (2016). Mobile manipulation, tool use, and intuitive interaction for cognitive service robot cosero. *Front. Robot. AI* 3:58. doi: 10.3389/frobt.2016.00058

ten Pas, A., Gualtieri, M., Saenko, K., and Platt, R. (2017). Grasp pose detection in point clouds. *Int. J. Robot. Res*. 36, 1455–1473. doi: 10.1177/0278364917735594

Thomas, G., Chien, M., Tamar, A., Ojea, J. A., and Abbeel, P. (2018). "Learning robotic assembly from cad," in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (Brisbane), 1–9.

Tosun, T., Daudelin, J., Jing, G., Kress-Gazit, H., Campbell, M., and Yim, M. (2018). "Perception-informed autonomous environment augmentation with modular robots," in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (Brisbane), 6818–6824.

Toth, N., Schick, K. D., Savage-Rumbaugh, E. S., Sevcik, R. A., and Rumbaugh, D. M. (1993). Pan the tool-maker: Investigations into the stone tool-making and tool-using capabilities of a bonobo (*Pan paniscus*). *J. Archaeol. Sci*. 20, 81–91.

Turner, M., Duggan, L., Glezerson, B., and Marshall, S. (2020). Thinking outside the (acrylic) box: a framework for the local use of custom-made medical devices. *Anaesthesia* 75, 1566–1569. doi: 10.1111/anae.15152

Wicaksono, H., and Sheh, C. S. R. (2017). "Towards explainable tool creation by a robot," in *IJCAI-17 Workshop on Explainable AI (XAI)* (Melbourne), 63.

Wohlkinger, W., and Vincze, M. (2011). "Ensemble of shape functions for 3d object classification," in *2011 IEEE International Conference on Robotics and Biomimetics (ROBIO)* (St. Paul, MN), 2987–2992.

Zech, P., and Piater, J. (2016). Grasp learning by sampling from demonstration. *arXiv [Preprint]*. arXiv:1611.06366. Available online at: https://arxiv.org/abs/1611.06366

Zhang, Z., Sturtevant, N. R., Holte, R. C., Schaeffer, J., and Felner, A. (2009). "A* search with inconsistent heuristics," in *IJCAI* (Pasadena, CA), 634–639.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

PERSPECTIVE
published: 23 March 2021
doi: 10.3389/frobt.2021.647028

Locating Creativity in Differing Approaches to Musical Robotics

*Steven Kemper**

Music Department, Mason Gross School of the Arts, Rutgers University, New Brunswick, NJ, United States

The field of musical robotics presents an interesting case study of the intersection between creativity and robotics. While the potential for machines to express creativity represents an important issue in the field of robotics and AI, this subject is especially relevant in the case of machines that replicate human activities that are traditionally associated with creativity, such as music making. There are several different approaches that fall under the broad category of musical robotics, and creativity is expressed differently based on the design and goals of each approach. By exploring elements of anthropomorphic form, capacity for sonic nuance, control, and musical output, this article evaluates the locus of creativity in six of the most prominent approaches to musical robots, including: 1) nonspecialized anthropomorphic robots that can play musical instruments, 2) specialized anthropomorphic robots that model the physical actions of human musicians, 3) semi-anthropomorphic robotic musicians, 4) non-anthropomorphic robotic instruments, 5) cooperative musical robots, and 6) individual actuators used for their own sound production capabilities.

Keywords: robotic musical instruments, creativity, anthropomorphism, music generation, musical robotics

OPEN ACCESS

Edited by:
Amy LaViers, University of Illinois at Urbana-Champaign, United States

Reviewed by:
Myounghoon Jeon, Virginia Tech, United States
Richard Savery, Georgia Institute of Technology, United States

Correspondence:
Steven Kemper
skemper@mgsa.rutgers.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *28 December 2020*
Accepted: *11 February 2021*
Published: *23 March 2021*

Citation:
Kemper S (2021) Locating Creativity in Differing Approaches to Musical Robotics. Front. Robot. AI 8:647028. doi: 10.3389/frobt.2021.647028

INTRODUCTION

The field of musical robotics presents an interesting case study of the intersection between creativity and robotics. While the potential for machines to express creativity represents an important issue in the field of robotics and AI, this subject is especially relevant in the case of machines that replicate human activities that are traditionally associated with creativity, such as music making. Several recent studies have explored the history and current state of musical robotics. While these present an overview of the field, they tend to focus primarily on issues related to functional design, with little discussion of creativity. Musical robots are categorized based on how they produce sound (Kapur 2005), how they function as interactive multimodal systems (Solis and Ng 2011), how they developed over history (Murphy et al., 2012; Long et al., 2017), and the ways that they engage in "Robotic Musicianship" (Bretan and Weinberg, 2016).[1]

Based on a review of existing literature as well as the author's experience designing and composing music for musical robots, this article proposes a new classification framework based on the ways that musical robots express creativity through anthropomorphic form, capacity for sonic nuance, control, and musical output. By exploring the field of musical robotics through this lens, we are able to better understand the ways that specific approaches lead to both technical and artistic goals.

[1]This article defines Robotic Musicianship as the intersection of musical mechatronics and machine musicianship. While the concepts of creativity and anthropomorphism are mentioned in passing, they are not used to classify different approaches to musical robotics.

DEFINITIONS AND CRITERIA FOR EVALUATION

Defining Musical Robotics

Both designers and audiences use the term "musical robots" or "robotic musical instruments" to refer to a broad range of musical machines. From an engineering perspective, approaches that lack autonomy could be more accurately be described as "musical mechatronics" (Bretan and Weinberg, 2016). However, the popular conception of robots, rooted in mythology, includes any machines that can mimic human actions (Jones 2017; Szollosy, 2017). Therefore, this discussion will consider "musical robotics" as any approach where an electromechanical actuator produces a visible, physical action that models the human act of music making, regardless of autonomous control.

By modeling the human act of music making, musical robots may be considered inherently anthropomorphic. Fink describes the important connection between anthropomorphism and robotics, as expressed through anthropomorphic form (appearance), behavior, and interaction with humans (Fink, 2012). While not all musical robots possess an anthropomorphic form, modeling the physical actions of music making represents anthropomorphic behavior. The ways that designers and audiences experience anthropomorphism significantly impacts how these machines express creativity. With this idea in mind, I identify six approaches that express creativity in different ways. These include: 1) nonspecialized anthropomorphic robots that can play musical instruments, 2) specialized anthropomorphic robots that model the physical actions of human musicians, 3) semi-anthropomorphic robotic musicians, 4) non-anthropomorphic robotic instruments, 5) cooperative musical robots, and 6) individual actuators used for their own sound production capabilities.

Defining Creativity

Several different fields currently focus on creativity, including esthetics, psychology, and artificial intelligence (Götz 1981; Bailin 1983; Boden, 1996; Boden, 2004; Cope 2005; Runco and Jaeger, 2012). Runco and Jaeger distinguish two fundamental criteria of creativity: originality and effectiveness (Runco and Jaeger 2012, 92). Originality, or creative insight, emerges from what Cope describes as, "The initialization of connections between two or more multifaceted things, ideas, or phenomenon hitherto not otherwise considered actively connected" (Cope 2005, 11). Effectiveness is determined through evaluation of creative insight by the creator, as well as related communities (Boden, 1996, 268).

Evaluative Criteria for Creativity in Musical Robotics

Musical robots tend to be viewed as creative machines due to their connection to music, which is understood to be an inherently creative endeavor. While studies of creativity in musical robots should focus on the music they produce, originality and effectiveness are also expressed through anthropomorphic form, capacity for sonic nuance, control, as well as musical output.

Anthropomorphic Form

The physical appearance of musical robots as well as the ways they model the human actions of music making are extremely important for designers and audiences. According to Fink, "the physical shape of a robot strongly influences how people perceive it and interact with it…" (Fink, 203). Fink also describes the importance of anthropomorphic behavior from the observer's perspective. "If a system behaves much like a human being (e.g., emits a human voice), people's mental model of the system's behavior may approach their mental model of humans," based on the estimation of the robot's capabilities (Fink, 201). Some approaches to musical robotics focus on modeling human appearance and movement while others explore mechatronic sound production techniques that do not possess anthropomorphic form. Evaluating creativity in terms of anthropomorphic form requires an understanding of the ways that designers and audiences ascribe human qualities to a musical robot's form and behavior.

Capacity for Sonic Nuance

Much of the existing literature in the field of musical robotics focuses on robots' ability to model the sonic capabilities of human performers. The benchmark for success in this area is often described as the ability to play music expressively (e.g., Murphy, 2014). While designers often describe how advancements in sound control parameters and their resolution enable expressivity, the concept of expression tends to be loosely defined (Kemper and Cypess, 2019). Therefore, it is more accurate to describe these features as increasing the capacity for sonic nuance (Kemper and Barton, 2018). While greater capacity for sonic nuance allows musical robots to more accurately model the dynamics, articulations, and phrasing of human performers, it can also create novel sonic and musical possibilities that differ from the ways that humans perform (Kemper, 2014). Thus, creativity in this domain refers to novel approaches to sonic nuance either for the purposes of modeling human performance or exploring new sonic and musical possibilities that are unique to musical robots.

Control

Musical robots can be controlled in a variety of ways, ranging from autonomous modes that enable interaction with human performers to modes where the movement of every actuator is preprogrammed. One of the challenges of assessing creativity in musical robotics is that control systems are often separable from the robot itself. Research in the areas of artificial musical generation and listening algorithms tend to focus on note generation in a generic way (e.g., as MIDI data), rather than being tailored to the mechanical requirements of a specific robot (e.g., Cope 2005; Xia and Dannenberg, 2015). For example, Solis and Ng's *Musical Robots and Interactive Multimodal Systems,* is divided into two separate sections that describe control and output respectively (Solis and Ng, 2011). While control determines the ways that actuators operate and thus how the robot produces sound, it is important to distinguish these instructions from the actual musical output.

Musical Output

The music that robots perform represents an important avenue for expressing creativity; however, this topic has received surprisingly little attention. Some robots use a single piece of music to demonstrate their capabilities, while others perform in a diverse array of styles, collaborate in real time with human performers, and are designed as creative tools for musical artists.[2] Evaluating creativity in musical output should consider both the robot's performative capabilities and musical decisions. Performative capabilities include the ability for robots to present a compelling performance, either by modeling human performers or exploring their own unique capabilities. Musical decisions include the specific musical pieces composed or arranged for the robot(s), musical decision-making by autonomous control systems, and the ways that new music created for (or by) robots engages with the unique capabilities of these machines.

DIFFERENT APPROACHES TO MUSICAL ROBOTICS

Nonspecialized Anthropomorphic Robots that Can Play Musical Instruments

Over the past 2 decades several companies have developed general-purpose anthropomorphic bipedal robots that replicate human actions in a variety of areas, including musical performance (Goswami and Vadakkepat 2019). For example, Toyota modified versions of their Partner robot to play trumpet, violin, and an electronic drum kit (Doi and Nakajima 2019). Of these approaches, the trumpet robot approximates human performance most closely in terms of articulation, dynamics, and timing. Conversely, the violin playing robot is limited in its range, and struggles somewhat with intonation and tone compared to a trained violinist.[3] This reflects the challenges of modeling the complex physical actions of bow pressure, bow speed, proper finger position, and vibrato.

In general, these demonstrations prioritize showing versatile, humanoid robots engaging in a quintessentially "human" activity over novel musical output. As Doi and Nakajima state, "We began the development of musical performance humanoid out of curiosity that we would like to make a humanoid robot realize such a human unique activity [*sic*]" (Doi and Nakajima, 218). This is emphasized by the fact that available videos of these robots perform easily recognizable versions of popular music, including "When you Wish Upon a Star" and "Pomp and Circumstance."[4,5] As robots are more specifically designed for musical performance, they become more specialized in their ability to produce sonic nuance as described in the examples below.

Specialized Anthropomorphic Robots That Model the Physical Actions of Human Musicians

Several approaches have focused on building robots that model the physical actions involved in musical performance. These include pioneering work from Waseda University, including the WABOT-series piano robot, WF-series flutist robot, and the WAS-series saxophone robot (Roads 1986; Solis et al., 2006; Solis and Hashimoto 2010). Shibuya and Park have also created robotic models of violin performance (Shibuya et al., 2007; Park et al., 2016), and Chadefaux has created a robotic "finger" for harp plucking (Chadefaux et al., 2012).

While these approaches accurately model the actions of human performance, they can result in a lack of musical "efficiency" when compared to musical robots that do not model human actions (see *Non-anthropomorphic Robotic Instruments*). For example, the Waseda WF-4RII Flutist Robot possesses 43 DOF, and each robotic component is designed to replicate its human counterpart, including "humanoid organs" such as robotic lips, lungs, arms, neck, tongue, and oral cavity (Solis et al., 2006, 13). By modeling the human actions of performance, the robot helps us to understand how instrumental performers produce musical sounds. However, the complexity of the mechanical model limits the sonic possibilities that are available to machines, such as super-virtuosic speed and novel approaches to sonic nuance. This is evidenced by available videos of performance, such as that of the WF-4RII performing Rimsky-Korsakov's "Flight of the Bumblebee" at a (humanly) comfortable tempo of c.150 BPM.[6]

Semi-Anthropomorphic Robotic Musicians

The musical robots in this category assume an anthropomorphic form, however they do not model the specific actions of human performance and are focused more on appearance and musical output. Over the past few decades several robotic "bands" have emerged, including the rock bands *The Trons, Captured! By Robots* and *Compressorhead*, as well as a collaboration between *Z-Machines*[7] and Squarepusher on the 2014 album "Music for Robots." (Snake-Beings, 2017; Gallagher, 2017; Davies and Crosby, 2016; Squarepusher x Z-Machines, 2014). *MOJA* features a drummer, harpist, and flutist performing in a style evocative of traditional Chinese music.[8] In addition to these "bands," the Robotic Musicianship Group at Georgia Tech has developed two well-documented semi-Anthropomorphic musical robots: Haile, a robotic drummer, and Shimon, a robotic marimba player (Weinberg and Driscoll 2006; Weinberg et al., 2020).

The anthropomorphic nature of these robots is highlighted primarily by their stylized appearance, rather than an attempt to

[2]E.g. https://www.patmetheny.com/orchestrioninfo/

[3]Toyota Partner Violin Robot: https://www.youtube.com/watch?v=-yInphJdick

[4]Toyota Partner Trumpet Robot: https://www.youtube.com/watch?v=6fctULDctuA

[5]Toyota Partner Violin Robot (see n.3).

[6]http://www.takanishi.mech.waseda.ac.jp/top/research/music/flute/wf_4rii/index.htm (Section IV)

[7]https://www.yurisuzuki.com/design-studio/z-machines

[8]https://news.tsinghua.edu.cn/en/info/1012/5231.htm

model the human actions of performance. For example, *Compressorhead*'s multi-armed drummer "Stickboy" has a mohawk made of metal spikes and is designed to headbang along with the music. *MOJA*'s robots are dressed in Tang-dynasty style garments. These design choices have nothing to do with sound production, however they enhance the connection between the audience and robot performers.

Even though the sound-producing mechanisms of these robots do not model human performers, much of the music they play could be easily performed by humans. One exception to this is Squarepusher's approach, which engages with the unique musical possibilities afforded by *Z-Machines's* robots. In the song "Sad Robot Goes Funny" the double-necked "guitar-bot" instrument performs extremely rapid picking while at the same time dynamically changing the chords in a way that would be impossible for a human musician. This takes full advantage of that instrument's 78 solenoid-based "fingers" and picks that can articulate each string individually. Similarly, while Shimon and Haile are designed to perform with human musicians, both have explored the extra-human musical capabilities of their designs with an emphasis on "play [ing] like a machine" (Weinberg et al., 2020, 95).

Non-Anthropomorphic Robotic Instruments

Non-anthropomorphic robotic instruments can either be mechatronic augmentations of existing acoustic instruments (e.g. Yamaha's Disklavier),[9] or newly designed instruments with no acoustic analog (e.g. Andy Cavatorta Studio's Gravity Harp).[10] **Table 1** includes a selection of recently active groups and individuals producing collections of non-anthropomorphic robotic instruments, as well as well-documented individual robots.

Non-anthropomorphic robotic instruments tend to focus more on sonic nuance than modeling the human actions of performance. For example, the Logos Foundation's robotic vibraphone <Vibi> couples actuating and dampening solenoids to each bar of the instrument rather than designing robotic arms and hands with multiple degrees of freedom that would model a human performer (Maes et al., 2011, 41).[11] This design allows <Vibi> to play much more rapidly than a human performer. It also enables complete polyphony of the instrument as well as individual control of the dampening mechanisms. <Vibi>'s unique capabilities open up a new world of musical possibilities when compared to a human performer on a traditional instrument.

One drawback of <Vibi>'s design is that since the solenoids are mounted below the striking bars it is difficult for the audience to see their movement, obscuring the connection between physical action and sound production. While this is a common issue in this category, some projects are designed to maximize the visibility of movement. For example, LEMUR's GuitarBot features four vertically mounted strings where pitch is changed on each string with a belt-driven fret that travels over half a meter (Singer et al., 2003). Other approaches include using LEDs to visualize sound production (e.g. Rogers et al., 2015).

Cooperative Musical Robots

An emerging area of musical robotics combines human performance and robotic actuation on a single shared interface. Barton describes these devices as cooperative musical machines, differentiating between cooperative (electro)mechanical instruments that do not react to human input, and cooperative robotic instruments that respond and interact with human performers (Barton, et al., 2017). Examples of cooperative (electro)mechanical instruments include Meywa Denki's *Ultra Folk* acoustic guitar[12] and Gurevich's *STRINGTREES* (Gurevich, 2014). Examples of cooperative robotic instruments include Barton's *Cyther*, a human-playable, self-tuning robotic zither, as well as the previously discussed Halie (Weinberg et al., 2020, 26).

Moving beyond a shared interface, Georgia Tech's Robotic Drumming Prosthetic Arm robotically augments the capabilities of the body. This device consists of a prosthetic arm outfitted with brushless gimbal motors and a single stage timing belt drive connected to a drumstick (Weinberg et al., 2020, 219). An amputee drummer controls the stick through EMG sensors connected to muscles on the residual limb. Rather than simply serving as a replacement for a human arm, the capabilities of mechatronic design and robotic control, including the addition of a second stick, allow for humanly impossible virtuosity and speed (Weinberg et al., 213, 226). Beyond musical possibilities, robotic augmentation of the body concretizes notions of posthumanism and the cyborg (Haraway 1991). It also causes observers to question the "humanness" of an augmented individual rather than evoking a sense of anthropomorphism (Swartz and Watermeyer 2008), though that may change as these technologies become more widely accepted.

Individual Actuators Used for Their Own Sound Production Capabilities

The final category in this discussion encompasses projects that focus on the sounds and movement of individual actuators. While some may not consider these approaches to be musical robots due to a lack of complexity in design or sonic output, I argue that they are important to consider in this discussion because 1) as with all of the other approaches described here, they possess electromechanical actuators that produce a visible, physical action resulting in sound production and 2) they distill sound produced by electromechanical actuators to its most basic form.

Several designers have created music using voice coil and stepper actuators from floppy disc and hard drives, as well as from individual stepper motors. These approaches tend to reproduce well-known music, such as Zadrożniak's arrangement of the "Imperial March" from *Star Wars* for floppy disc drives.[13] While these actuators produce a unique timbre, there is limited capacity for sonic nuance.

Other designers create work that produces sound using the simple actions of motors. For example, Zimoun builds large-scale sound sculptures that feature individual motors actuating resonant objects.[14] One installation consists of 658 cardboard

[9]https://usa.yamaha.com/products/musical_instruments/pianos/disklavier/index.html

[10]https://andycavatorta.com/gravityharps.html

[11]https://logosfoundation.org/instrum_gwr/vibi.html

[12]https://www.maywadenki.com/sketch/tsukuba-2/

[13]https://www.youtube.com/watch?v=yHJOz_y9rZE

[14]https://www.zimoun.net/

TABLE 1 | Selection of recently active groups and individuals producing collections of non-anthropomorphic robotic instruments, as well as recently designed, well-documented individual robots.

Group name	Website	Citation
Andy Cavatorta Studio	https://andycavatorta.com/	n/a
EMMI: Expressive Machines Musical Instruments	https://www.youtube.com/user/ExpressiveMachines	Rogers et al. (2015)
Karmetic Machine Orchestra	https://www.karmetik.com/karmetik-projects/the-machine-orchestra/	Kapur et al. (2011)
LEMUR: League of Electronic Musical Urban Robots	http://lemurbots.org/index.html	Singer et al. (2004)
Logos Foundation Man and Machine Orchestra	https://www.logosfoundation.org/mnm/index.html	Maes et al. (2011)
Matthew Steinke	https://www.matthewsteinke.com/robotic	n/a
Maywa Denki	https://www.maywadenki.com/sketch/tsukuba-2/	Ippolito (2009)
Moritz Simon Geist	http://sonicrobots.com/	n/a
Music, Perception and Robotics Lab	https://mprlab.org/	Barton et al. (2018)
Robot Rickshaw	http://www.troy82.com/musical-robots/projects/robot-rickshaw/	n/a
Sonic Engineering Lab for Creative Technology	https://ecs.wgtn.ac.nz/Groups/SELCT/IndexProto	Murphy (2014)
Studio Jose de la O	http://delao.mx/robokumbia	Flores et al. (2019)
Trimpin	https://www.historylink.org/File/11204	Leitman (2011)

Individual Robots	Website	Citation
Berimbot	http://www.berimbo.com.br/	Monsão et al. (2017)
Gamelatron	https://gamelatron.com/	McGraw (2016)
McBlare	https://www.cs.cmu.edu/~music/mcblare/	Dannenberg et al. (2011)
MechDrum	https://mechdrum.wordpress.com/	Van Rooyen et al. (2017)

boxes that are hit by cotton balls connected to a DC motor by piano wire. As the motor spins the ball hits the box and produces a resonant sound. In this approach, gravity, friction, and resonance, as well as the movement of the motor itself produce variations in the sound that makes the work compelling.

DISCUSSION

This paper proposes a novel classification system that enables us to consider how different approaches to musical robotics express creativity in different ways. In general, the more overtly anthropomorphic the form, the more central the physical appearance of the project is to its creativity. For example, the Toyota Partner robot's creative impact stems from the fact that a humanoid robot is performing a quintessentially human activity. As anthropomorphic form diminishes, originality and effectiveness are conveyed through the capacity for sonic nuance, as well as the ways that these machines either accurately model human performance or develop their own robotic performance practice. For an extreme case such as Zimoun's work, the connection to anthropomorphic behavior in the process of sound production lies at the center of the work. By understanding the connections and divergent goals among different approaches to musical robotics, we can better evaluate the ways that these machines express originality and effectiveness for both designers and audiences. Though the classifications developed here are theoretical in nature, they will hopefully prove useful in developing future studies that explore the ways that musical machines can be considered creative.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Bailin, S. (1983). On creativity as making: a reply to götz. *J. Aesthetics Art Criticism* 41 (4), 437–442. doi:10.2307/429877

Barton, S., Prihar, E., and Carvalho, P. (2017). "Cyther: a human-playable, self-tuning robotic zither," in Proceedings of the international conference on new interfaces for musical expression, Copenhagen, Denmark (NIME), 319–324.

Barton, S., Sundberg, K., Walter, A., Baker, L.S., Sane, T., and O'Brien, A. (2018). "A robotic percussive aerophone," in *Proceedings of the 18th international conference on new interfaces for musical expression*. 409–412.

Boden, M. A. (1996). "Creativity," in *Artificial intelligence*. Editor M. A. Boden (San Diego, CA: Academic Press), 267–291.

Boden, M. A. (2004). *The creative mind: myths & mechanisms*. 2nd Edn. London, United Kingdom: Routledge.

Bretan, M., and Weinberg, G. (2016). A survey of robotic musicianship. *Commun. ACM* 59 (5), 100–109. doi:10.1145/2818994

Chadefaux, D., Le Carrou, J-L., Vitrani, M-A., Billout, S., and Quartier, L. (2012). "Harp plucking robotic finger," in 2012 IEEE/RSJ international conference on intelligent robots and systems (Vilamoura-Algarve, Portugal), 4886–4891. doi:10.1109/IROS.2012.6385720

Cope, D. (2005). *Computer models of musical creativity*. Cambridge, MA: MIT Press.

Dannenberg, R. B., Brown, H. B., and Lupish, R. (2011). "McBlare: a robotic bagpipe player," in *Musical robots and interactive multimodal systems. Springer tracts in advanced robotics*. Editors J. Solis and K. Ng (Berlin: Springer).

Davies, A., and Crosby, A. (2016). "Compressorhead: the robot band and its transmedia storyworld," in *Cultural robotics*. Editors J. T. K. V. Koh, B. J. Dunstan, D. Silvera-Tawil, and M. Velonaki (Cham: Springer International Publishing).

Doi, M., and Nakajima, Y. (2019). "Toyota partner robots," in *Humanoid robotics: a reference*. Editors A. Goswami and P. Vadakkepat (Dordrecht: Springer Nature).

Fink, J. (2012). "Anthropomorphism and human likeness in the design of robots and human-robot interaction," in *Social robotics*. Editors S. Sam Ge, O. Khatib, J. Cabibihan, R. Simmons, and M-A. Williams (Berlin, Heidelberg: Springer), 199–208.

Flores, R. I., Morán, R. M. L., and Ruano, D. S. (2019). Mexi-futurism. The transitorial path between tradition and innovation. *Strateg. Des. Res. J.* 12 (2), 222–234. doi:10.4013/sdrj.2019.122.08

Gallagher, D. (2017). Jay vance may play metal with homemade robots, but He says it's not meant to be funny. Available at: https://www.dallasobserver.com/music/jay-vance-frontman-of-captured-by-robots-on-how-he-turned-his-novelty-robot-band-into-a-serious-one-10081744 (Accessed December 10, 2020).

A. Goswami and P. Vadakkepat (2019). *Humanoid robotics: a reference* (Dordrecht, Netherlands: Springer).

Götz, I. L. (1981). On defining creativity. *J. Aesthetics Art Criticism* 39 (3), 297–301. doi:10.2307/430164

Gurevich, M. (2014). "Distributed control in a mechatronic musical instrument," in Proceedings of the international conference on new interfaces for musical expression, 487–490. London, United Kingdom: Zenodo.

Haraway, D. (1991). *"A cyborg manifesto: science, Technology, and socialist-feminism in the late twentieth century" in simians, cyborgs, and women: the reinvention of nature*. New York, NY: Routledge.

Ippolito, J. (2009). Art commodities from Japan: propagating art and culture via the internet. *Int. J. Arts Soc.* 4 (4), 109–116. doi:10.18848/1833-1866/cgp/v04i04/35682

Jones, R. (2017). Archaic man meets a marvellous automaton: posthumanism, social robots, archetypes. *J. Anal. Psychol.* 62, 338–355. doi:10.1111/1468-5922.12316

Kapur, A. (2005). "A history of robotic musical instruments," in Proceedings of the 2005 international computer music conference, ICMC 2005. September 4-10, 2005, Barcelona, Spain. Available at: http://hdl.handle.net/2027/spo.bbp2372.2005.162.

Kapur, A., Darling, M., Diakopoulos, D., Murphy, J. W., Hochenbaum, J., Vallis, O., et al. (2011). The machine orchestra: an ensemble of human laptop performers and robotic musical instruments. *Comp. Music J.* 35 (4), 49–63. doi:10.1162/comj_a_00090

Kemper, S., and Barton, S. (2018). "Mechatronic expression: reconsidering expressivity in music for robotic instruments," in Proceedings of the 18th international conference on new interfaces for musical expression. Blacksburg, Virginia, USA: Zenodo, 84–87. doi:10.5281/zenodo.1302689

Kemper, S. (2014). Composing for musical robots: aesthetics of electromechanical music. *Emille: J. Korean Electro-Acoustic Music Soc.* 12, 25–31.

Kemper, S., and Cypess, R. (2019). Can musical machines Be expressive? Views from the enlightenment and today. *Leonardo* 52 (5), 448–454. doi:10.1162/LEON_a_01477

Leitman, S. (2011). Trimpin: an interview. *Comp. Music J.* 35 (4), 12–27. doi:10.1162/COMJ_a_00088

Long, J., Murphy, J., Carnegie, D., and Kapur, A. (2017). Loudspeakers Optional: a history of non-loudspeaker-based electroacoustic music. *Organised Sound* 22 (2), 195–205. doi:10.1017/S1355771817000103

Maes, L., Raes, G-W., and Rogers, T. (2011). The man and machine robot orchestra at Logos. *Comp. Music J.* 35 (4), 28–48. doi:10.1162/comj_a_00089

McGraw, A. (2016). Atmosphere as a concept for ethnomusicology: comparing the gamelatron and gamelan. *Ethnomusicology* 60 (1), 125–147. doi:10.5406/ethnomusicology.60.1.0125

Monsão, I. C., Cerqueira, J. D. J. F., and da Costa, A. C. P. L. (2017). The berimbot: a robotic musical instrument as an outreach tool for the popularization of science and technology. *Int. J. Soc. Robotics* 9, 251–263. doi:10.1007/s12369-016-0386-3

Murphy, J. (2014). Expressive musical robots: building, evaluating, and interfacing with an ensemble of mechatronic instruments. PhD dissertation. Wellington, NZ: Victoria University of Wellington.

Murphy, J., Kapur, A., and Carnegie, D. (2012). Musical robotics in a loudspeaker world: developments in alternative approaches to localization and spatialization. *Leonardo Music J.* 22, 41–48. doi:10.1162/lmj_a_00090

Park, H., Lee, B., and Kim, D. (2016). Development of anthropomorphic robot finger for violin fingering. *ETRI J.* 38, 1218–1228. doi:10.4218/etrij.16.0116.0129

Roads, C. (1986). The tsukuba musical robot. *Comp. Music J.* 10 (2), 39–43. doi:10.2307/3679483

Rogers, T., Kemper, S., and Barton, S. (2015). "MARIE: monochord-aerophone robotic instrument ensemble," in Proceedings of the international conference on new interfaces for musical expression. Baton Rouge, Louisiana, USA: Zenodo, 408–411. doi:10.5281/zenodo.1179166

Runco, M. A., and Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Res. J.* 24 (1), 92–96. doi:10.1080/10400419.2012.650092

Shibuya, K., Matsuda, S., and Takahara, A. (2007). "Toward developing a violin playing robot - bowing by anthropomorphic robot arm and sound analysis," in RO-MAN 2007 - the 16th IEEE international symposium on robot and human interactive communication. Jeju, Korea (South), 763–768. doi:10.1109/ROMAN.2007.4415188

Singer, E., Feddersen, J., Redmon, C., and Bowen, B. (2004). "LEMUR's musical robots," in Proceedings of the international conference on new interfaces for musical expression. Hamamatsu, Japan: Zenodo. doi:10.5281/zenodo.1176669

Singer, E., Larke, K., and Bianciardi, D. (2003). "LEMUR GuitarBot: MIDI robotic string instrument," in Proceedings of the international conference on new interfaces for musical expression, 188–191.

Snake-Beings, E. (2017). The do-it-yourself (DiY) craft aesthetic of the trons – robot garage band. *Craft Res.* 8 (1), 55–77. doi:10.1386/crre.8.1.55_1

Solis, J., Chida, K., Taniguchi, K., Hashimoto, S. M., Suefuji, K., and Takanishi, A. (2006). The Waseda flutist robot WF-4RII in comparison with a professional flutist. *Comp. Music J.* 30 (4), 12–27. doi:10.1162/comj.2006.30.4.12

Solis, J., and Hashimoto, A. K. (2010). "Development of an anthropomorphic saxophone-playing robot," in *Brain, body and machine proceedings of an international symposium on the occasion of the 25th anniversary of McGill university centre for intelligent machines*. Editors B. Boulet, J. J. Clark, J. Kovecses, and K. Siddiqi (Berlin: Springer), 175–186. doi:10.1007/978-3-642-16259-6

J. Solis and K. Ng (2011). *Musical robots and interactive multimodal systems* (Berlin: Springer).

Squarepusher x Z-Machines (2014). Music for robots. *Warp Rec.*

Swartz, L., and Watermeyer, B. (2008). Cyborg anxiety: oscar Pistorius and the boundaries of what it means to be human. *Disabil. Soc.* 23 (2), 187–190. doi:10.1080/09687590701841232

Szollosy, M. (2017). Freud, Frankenstein and our fear of robots: projection in our cultural perception of technology. *AI Soc.* 32, 433–439. doi:10.1007/s00146-016-0654-7

Van Rooyen, R., Schloss, A., and Tzanetakis, G. (2017). Voice coil actuators for percussion robotics. doi:10.5281/zenodo.1176149

Weinberg, G., and Driscoll, S. (2006). Toward robotic musicianship. *Comp. Music J.* 30 (4), 28–45. doi:10.1162/comj.2006.30.4.28

Weinberg, G., Bretan, M., Hoffman, G., and Driscoll, S. (2020). *Robotic musicianship: embodied artificial creativity and mechatronic musical expression*. Switzerland: Springer Nature.

Xia, G., and Dannenberg, R. B. (2015). "Duet interaction: learning musicianship for automatic accompaniment," in Proceedings of the international conference on new interfaces for musical expression. Baton Rouge, Louisiana, USA: Zenodo, 259–264. doi:10.5281/zenodo.1179198

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 29 April 2021
doi: 10.3389/frobt.2021.662355

Check for updates

Before, Between, and After: Enriching Robot Communication Surrounding Collaborative Creative Activities

Richard Savery, Lisa Zahray and Gil Weinberg*

Robotic Musicianship Lab, Georgia Tech Center for Music Technology, Atlanta, GA, United States

Research in creative robotics continues to expand across all creative domains, including art, music and language. Creative robots are primarily designed to be task specific, with limited research into the implications of their design outside their core task. In the case of a musical robot, this includes when a human sees and interacts with the robot before and after the performance, as well as in between pieces. These non-musical interaction tasks such as the presence of a robot during musical equipment set up, play a key role in the human perception of the robot however have received only limited attention. In this paper, we describe a new audio system using emotional musical prosody, designed to match the creative process of a musical robot for use before, between and after musical performances. Our generation system relies on the creation of a custom dataset for musical prosody. This system is designed foremost to operate in real time and allow rapid generation and dialogue exchange between human and robot. For this reason, the system combines symbolic deep learning through a Conditional Convolution Variational Auto-encoder, with an emotion-tagged audio sampler. We then compare this to a SOTA text-to-speech system in our robotic platform, Shimon the marimba player.We conducted a between-groups study with 100 participants watching a musician interact for 30 s with Shimon. We were able to increase user ratings for the key creativity metrics; novelty and coherence, while maintaining ratings for expressivity across each implementation. Our results also indicated that by communicating in a form that relates to the robot's core functionality, we can raise likeability and perceived intelligence, while not altering animacy or anthropomorphism. These findings indicate the variation that can occur in the perception of a robot based on interactions surrounding a performance, such as initial meetings and spaces between pieces, in addition to the core creative algorithms.

Keywords: creativity, robotics, music, improvisation, sound, text-to-speech, human-robot interaction

Edited by:
Patrícia Alves-Oliveira, University of Washington, United States

Reviewed by:
Nikolas Martelaro, Carnegie Mellon University, United States
Naomi Talya Fitter, University of Southern California, Los Angeles, United States

****Correspondence:***
Richard Savery
rsavery3@gatech.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *01 February 2021*
Accepted: *12 April 2021*
Published: *29 April 2021*

Citation:
Savery R, Zahray L and Weinberg G (2021) Before, Between, and After: Enriching Robot Communication Surrounding Collaborative Creative Activities. Front. Robot. AI 8:662355. doi: 10.3389/frobt.2021.662355

1 INTRODUCTION

There is a growing body of work focusing on robots collaborating with humans on creative tasks such as art, language, and music. The development of robotic functionalities leading to and following after collaborative creative tasks has received considerably less attention. These functionalities can address, for example, how a robot communicates and interacts with collaborators between musical improvisations, or before a piece begins or ends. Embodying a creative robot with speech capabilities that do not specifically address its creative capabilities risks distancing collaborators and misrepresenting artistic opportunities. In robotic literature this is referred to as the habitability gap, which addresses the problematic distance between a robot's implied capabilities and its actual potential output (Moore, 2017). In addition, human-robot collaboration is dependent on the development of a relationship between human and robot (Fischer, 2019). Emotion and personality conveyance has been shown to enhance robotic collaborations, with improved human-robot relationships and increased trust (Bates, 1994). One under-explored approach for an artificial agent to convey emotions is through non-linguistic musical prosody (Savery et al., 2020a). We propose that such an approach could be particularly effective in human-robot collaboration in creative tasks, where emotional expression is at the core of the activity, and where subtle background conveyance of mood can enhance, rather than distract, from the creative activity.

We present a model for generating emotional musical prosody in embedded platforms in real time for creative robots. The system aims to address the habitability gap by enriching human-robot communication before, during and after collaborative creative interaction. To support the system, we have created a new dataset of improvised emotional sung phrases, used to generate new emotional midi phrases through a convolutional variational autoencoder (CVAE) conditioned on emotion.

We implement this system in a marimba playing robot, Shimon, and analyze the impact on users during creativity-based musical interactions. The musical tasks feature call and response musical improvisation over a pre-recorded playback. We compare the perception of common metrics of likeability and perceived intelligence, with the perceived creativity and preferences for interaction as well as Boden's creativity metrics (Boden, 2009). We demonstrate that by using a creative communication method in addition to the core creative algorithms of a robotic system we are able to improve the interaction based on these metrics. Our implementation leads to the perception of higher levels of creativity in the robot, increased likeability, and improved perceived intelligence.

2 RELATED WORK

2.1 Human-Robot Communication

Verbal language-based interaction is the prominent form of communication used in human-robot interaction (Mavridis, 2015) covering a wide range of tasks from robot companions (Dautenhahn et al., 2006) to industrial robots (Pires and Azar, 2018). Many robotic interactions do not include language; these non-verbal forms of communication fall into six categories: kinesics, proxemics, haptics, chronemics, vocalics, and presentation (Jones, 2013; Saunderson and Nejat, 2019). Kinesics includes communication through body movement, such as gestures (Gleeson et al., 2013), or facial expressions, while proxemics focuses on the robotic positioning in space, such as the distance from a human collaborator (Walters et al., 2005). Haptics refers to touch based methods (Fukuda et al., 2012), while chronemics includes subtle traits such as hesitation (Moon et al., 2011). Presentation includes the way the robot appears, such as changes based on different behavior (Goetz et al., 2003). The final category, vocalics, includes concepts such as prosody (Crumpton and Bethel, 2016), which have shown to improve trust and other human-robot interaction metrics (Savery et al., 2019a). The vast majority of these communication techniques require significant technical and financial expense and variation to a system, such as adding augmented reality technology or changing robot movements (Saunderson and Nejat, 2019). In comparison, musical prosody can be implemented in an existing system with only minor changes (Savery et al., 2019b).

2.2 Musical Generation

Music generation has been widely addressed as a deep learning task (Briot et al., 2017), in particular using LSTMs (Sturm et al., 2016; Wu et al., 2019) and more recently transformers Huang et al. (2018). Music tagged with emotion has also been generated through long short-term memory networks (LSTMs) with logistic regression and used to generate music with sentiment (Ferreira and Whitehead, 2019). Other efforts have used a Biaxial LSTM network (Zhao et al., 2019), generating symbolic polyphonic musical phrases corresponding to Russel's valence-arousal emotion space (Posner et al., 2005). Variational autoencoders (VAEs) Kingma and Welling (2013); Rezende et al. (2014) use an encoder to represent its input probabilistically in latent space, and a decoder to convert from latent space back to the original input. Such VAEs have seen recent success in music generation tasks, for example, MIDI-VAE which use a VAE with recurrent encoder/decoder pairs to perform style transfer on midi data, changing the genre or composer of a piece (Brunner et al., 2018). MusicVAE employs a hierarchical decoder to better represent the long-term structure present in music, generating midi phrases that were 16 bars (about 30 s) long (Roberts et al., 2018).

3 CUSTOM DATASET

For this project we created a custom dataset of 4.22 h of audio recorded by Mary Esther Carter[1]. Carter is a professional vocalist and improviser who the authors have worked with before and were confident would be able to create a dataset matching the projects goals. Before collecting the data, we conducted exploratory sessions with seven different student musicians, comparing their ability to improvise different emotions using different classification systems. We additionally evaluated how well the musicians in this group could recognize the emotions played by other musicians. This process consisted of a 45 min in-person session, with musicians first improvising, followed by an informal interview to discuss the difficulty and their preferences for emotional classifications for improvisation. After these sessions, we decided that the Geneva Emotion Wheel (GEW) (Sacharin et al., 2012) was best suited for our purposes. The GEW is a circular model, containing 20 emotions with emotions and position corresponding to the circumplex model.

Our decision to use the GEW was based on multiple factors, firstly we aimed to capture as large a range of emotions as

[1]https://maryesthercarter.com/

possible, that could be accurately improvised by musicians in the sessions. In our exploratory study, the GEW balanced between having many recognizable classes, while also avoiding the potential confusion from too many overlapping classes, or the challenge of continuous classes such as the circumplex model. The GEW also has advantages for implementation, with 20 different discrete emotions which can be reduced to four separate classes, aligned with a quadrant from the circumplex model. GEW also includes most of the Eckman's basic emotions—fear, anger, disgust, sadness, happiness—only leaving out surprise. The ability to potentially reduce our collected dataset between these different models of emotion allows for significant future use cases.

It should be noted that this dataset comes from only one musician, and therefore only captures one perspective on musical emotion. While the dataset can make no claim to represent cross-cultural emotion conveyance and does not create a generalized emotion model, we believe that only collecting data from one person has advantages. By having only one vocalist our system can recreate one person's emotional style, avoiding incorrectly aggregating multiple styles to remove distinctive individual and stylistic features.

3.1 Process and Data

We first created a short list of vocalists who we have worked with in the past. We then conducted Skype calls with three professional vocalists refining the overall plan and describing the process, before asking Mary Carter to record and emotionally label her vocal improvisation. We choose to work with Carter as she had at home access to high quality recording equipment, and the authors have previously worked with her. In the future we expect to record with additional vocalists. Carter was paid $500 to record the samples over a week long period at her home studio, using a template we created in Apple digital audio workstation—Logic Pro, while maintaining the same microphone positioning. For the samples we requested phrases to be between 1 and 20 s, and to spend about 15 min on each emotion, allowing unscripted jumping between any order of the emotions. We allowed deletion of a phrase if the singer felt retroactively that the phrase did not capture the correct emotion. The final recorded dataset includes 2,441 phrases equaling 4.22 h of data with an average of 122 phrases for each emotion. Samples from the dataset can be heard online.[2]

3.2 Dataset Validation

To validate the dataset, we performed a study with 45 participants from Prolific and Mechanical Turk, paying each $3. Each question in the survey asked the participant to listen to a phrase and select a location on the wheel corresponding to the emotion and intensity they believed the phrase was trying to convey. Phrases fell under two categories of "best" and "all," with each participant listening to 60 total phrases selected at random. Between the 45 participants listening to 60 phrases, 2,700 ratings were given, which we believe gave a strong overall rating of the dataset. The "best" category consisted of five phrases for each emotion that were hand-selected by the authors as best representing that emotion. The best emotions were chosen to ensure an even distribution of phrase lengths in each emotion set, with each emotion having a chosen phrase for the lengths, 3, 5, 7, 9, and 11 s. When multiple phrases existed for each length the authors chose phrases that were most distinctive in style from the other emotions, aiming to create a stylistic separation between each emotion class. The "all" category consisted of a phrase sampled from all phrases in the dataset for that emotion, with a new phrase randomly selected for each participant. Rose plots of the validation results that combine the "best" and "all" categories can be seen in **Figure 1**, separated into each Geneva Wheel quadrant. The plots show strong validation correlation in Quadrants 1, 2 and 3, while Quadrant four showed a higher level of confusion.

3.3 Dataset to Midi

We converted each phrase's audio into a midi representation to use as training data. This process required significant iteration, as we developed a custom pipeline for processing our dataset. This was necessary due to the range of vocal timbre and effect, ranging from clear melodies, to non-pitched effects. We first ran the monophonic pitch detection algorithm CREPE (Kim et al., 2018) on each phrase, which output a frequency and a confidence value for a pitch being present every 0.01 s. As the phrases included breaths and silence, it was necessary to filter out pitches detected with low confidence. We applied a threshold followed by a median filter to the confidence values, and forced each detected pitch region to be at least 0.04 s long.

We next converted the frequencies to midi pitches. We found the most common pitch deviation for each phrase using a histogram of deviations, shifting the midi pitches by this deviation to tune each phrase. We rated onsets timing confidence between 0 and 1. To address glissando, vibrato and other continuous pitch changes, we identified peaks in the absolute value of the pitch derivative, counting an onset only when detecting a pitch for at least 0.04 s.

3.4 Scales

Scherer has shown that musical scales—without a melody or rhythm - are able to display emotion (Scherer et al., 2017). We therefore asked the singer to also record scales tagged with emotion to be used in an audio sampler. The audio sampler was designed to play back each note from the recorded scales, in such a way that new symbolic phrases consist of combinations of each note from the scale. In contrast to the main dataset we only recorded scales for four emotion classes, corresponding with each quadrant of the circumplex model. In addition to explaining the model to the vocalist, each quadrant had two key words which were angry/anxious, happy/exciting, relaxing/serene, sad/bored.

The data collection plan was based around common practice described by virtual instrument libraries[3]. For each emotion, 11 versions of a chromatic scale across an octave and a half were

[2]www.richardsavery.com/prosodycvae

[3]https://www.spitfireaudio.com/editorial/in-depth/grow-your-own-samples/

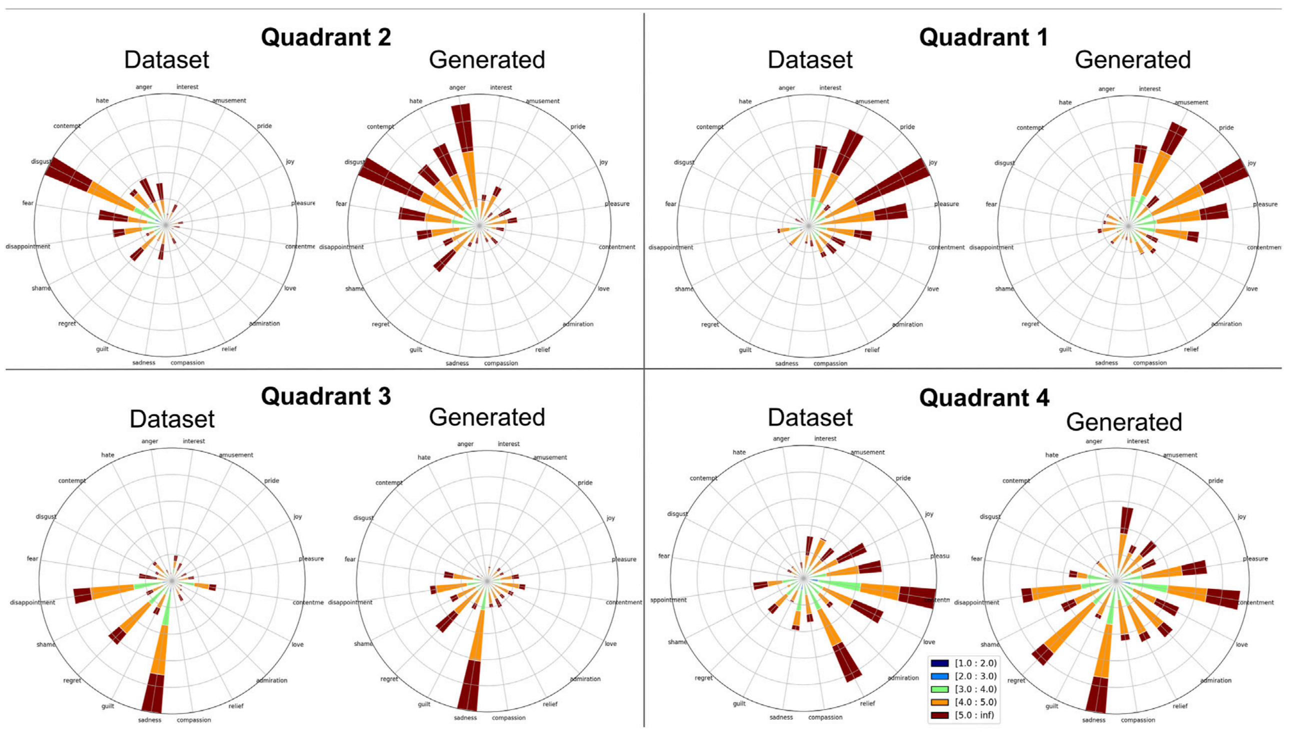


FIGURE 1 | Rose plots of dataset validation and generation evaluation.

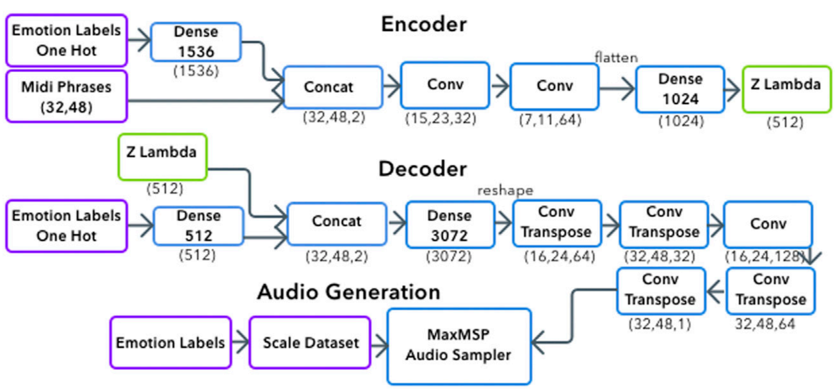


FIGURE 2 | Generative System Overview.

sung, 3 with very short notes, 3 with 500 ms, 3 with 1000 ms and 2 with 2000 ms duration. To allow the scales to contain all timbrel features for each emotion, we allowed for any dynamic variations and accents. The syllables themselves were also chosen for each scale by the vocalist.

4 GENERATIVE SYSTEM DESIGN

The system was designed with the primary goal of operating and responding to audio in real time on multiple embedded platforms. Future use cases will likely involve other computationally expensive systems, such as speech recognition and emotional interactions. In past work we have generated raw-audio for prosody (Savery et al., 2019b), however even after considerable refinement, and the use of multi-GPU systems, generation required 3 s of processing per 1 s of audio. With this in mind the initial design choice was to generate symbolic data using a version of the dataset converted to midi values, and not attempt to generate raw audio.

The symbolic generation of the system contains the pitch and rhythm of emotionally labeled melodies. Due to the process

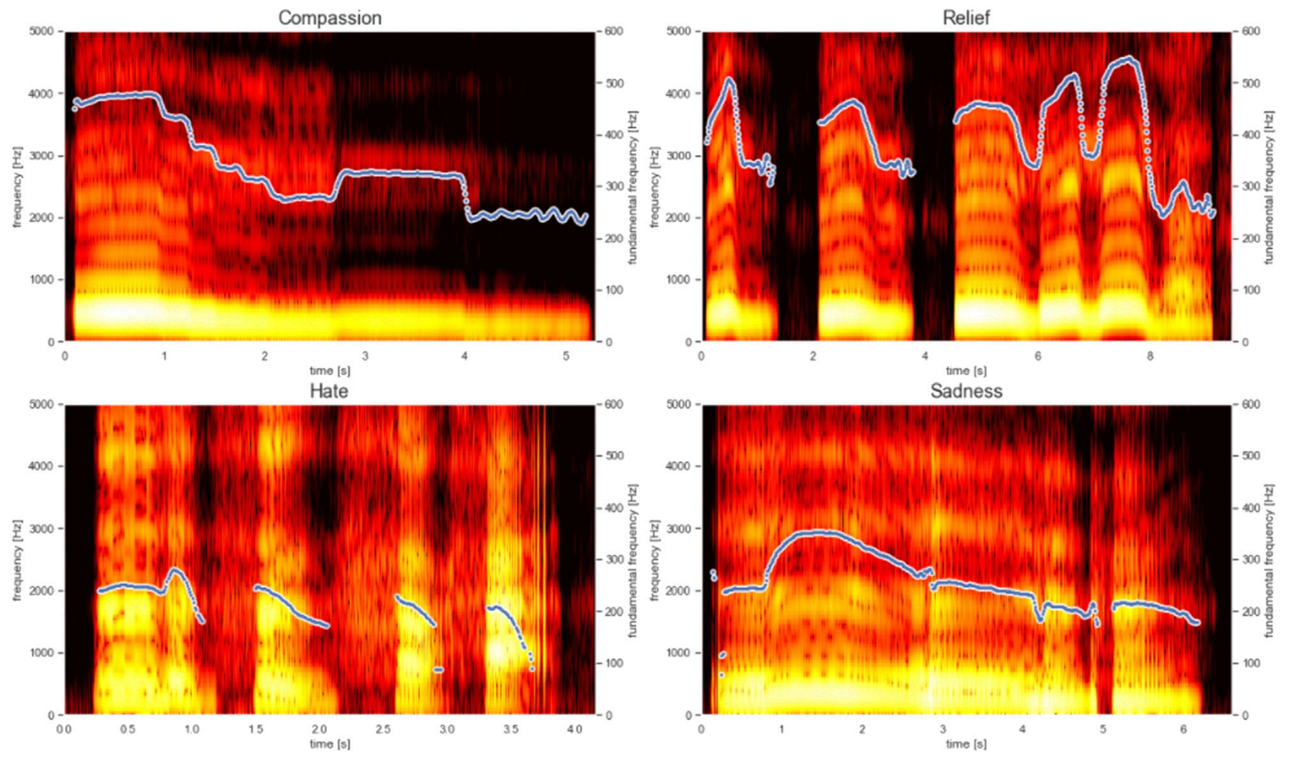


FIGURE 3 | Spectogram of musical prosody phrases (blue line indicates pitch contour).

described in **Section 3.3** the data also includes micro-timings. Symbolic data alone does not capture the range of emotion present in the dataset through timbre variations. By using the scale dataset described in **Section 3.4** the generation process encapsulates symbolic information with tagged emotion, capturing timbre and phoneme information. **Figure 2** shows an overview of the system. The system's interface is written MaxMSP, allowing users to chose an emotion. This activates a python script which generates a midi file and returns it to MaxMSP. **Figure 3** presents an example of the musical prosody phrases the systems is capturing, showing the contrasting pitch, rhythm and timbre for each emotion. Generated samples can be heard online.[4]

4.1 CC-VAE

4.1.1 Data Representation

We maintain the same data structure as developed in our audio to midi process, using midi pitch values that are sampled every 10 milliseconds. We then convert each melody to a length of 1,536 samples, and zero pad shorter melodies. Versions of each phrase are then transposed up and down six semitones, to give 12 versions of each phrase, one in each key. The melody is then reshaped to be 32 by 48 samples. The emotion label for each melody is converted to a one-hot representation.

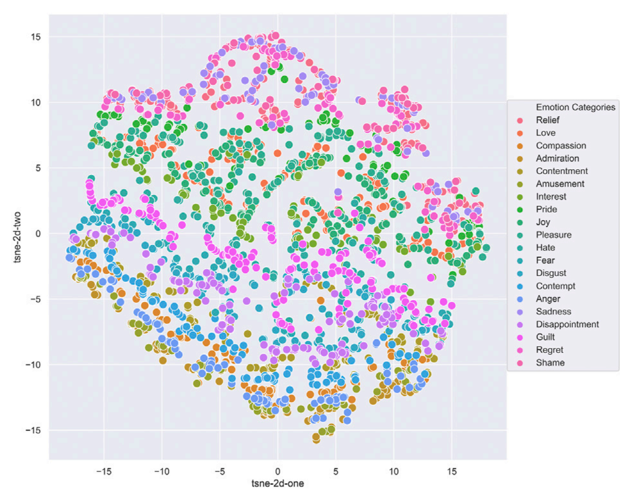


FIGURE 4 | Vanilla VAE Latent Space, classifying Carter's audio dataset.

4.1.2 Network Design

We chose to use VAEs due to their recent success in sequence and music generation tasks, and because they allow for analysis of the latent space which can provide insight into how well the network has learned to represent the different emotions. VAEs can be used to generate new data by sampling and decoding from the latent space, allowing the system to learn features of the data in an unsupervised manner. **Figure 4** shows the latent space after

[4]www.richardsavery.com/prosodycvae

TABLE 1 | Results of emotion survey for dataset phrases compared with generated phrases. See *Generation Evaluation* for an explanation of the metrics.

Quadrant	% Correct Quadrant		Average Difference		Average Variance	
	Dataset	Generated	Dataset	Generated	Dataset	Generated
1	57.2	56.3	1.32	1.98	1.76	1.83
2	54.5	52.5	1.45	0.96	1.79	1.88
3	57.4	51.5	2.16	1.93	1.92	1.89
4	43.7	31.9	1.61	1.24	1.86	2.03

training a Vanilla VAE on our custom dataset, without emotion labels. This demonstrates the latent space is able to separate by emotion without conditioning.

Our Conditional VAE is based on the standard architecture proposed by Sohn et al. (Sohn et al., 2015). A Conditional Variational Encoder (CVAE) varies from a VAE by allowing an extra input to the encoder and decoder. We input a one-hot emotion label, allowing for sampling a specific emotion from the latent space. As is common practice for a VAE, we use Kullback-Leibler divergence in the loss function. Our latent space dimension is 512, which we arrived at after testing multiple variations.

We chose to use a Convolutional Network (ConvNet) within our CVAE for multiple reasons. Although ConvNets are much less common in symbolic music generation (Briot et al., 2017), they have been used for audio generation such as WaveNet (Oord et al., 2016) as well as some symbolic generations (Yang et al., 2017). While we experimented with Vanilla RNNs, LSTMs and GRUs as encoders and decoders we found they were very prone to overfitting when trained conditionally, likely due to our dataset size. Our architecture is presented in **Figure 2**.

4.2 Sample Player

The generated midi file is loaded into MaxMSP to be played by the sampler. The audio sampler plays back individual notes created during the recording of the scales. MaxMSP parses the midi file, assigning each note a midi channel. Channels are divided by emotion and note length. For example, happy is assigned to channels one to four, with channel one containing the shortest note and channel four the longest note; sad is assigned to channels five to eight with the shortest note assigned to channel five and the longest note assigned to channel 8. The audio sampler plays as a midi device, and can be played directly like any midi instrument.

4.3 Generation Evaluation

To evaluate the results, we first generated three phrases for each emotion. We then ran a survey using the same questions as the dataset validation described in **Section 3.2**, asking 39 new participants to select an emotion and intensity for each of the 60 total generated phrases. Participants encountered five listening tests during the survey, and we only used data from participants who answered all listening tests correctly. **Figure 1** shows a comparison between the rose plots for each quadrant of the original dataset vs. the generated phrases.

We computed the mean and variance for each emotion, weighted by intensity, using the methods described in (Coyne et al., 2020), which rely on circular statistics. The results are shown in **Table 1**. The first columns show the percentage of all data points that were classified as an emotion in the correct quadrant. The next columns, showing average difference, were calculated by first finding the difference between each ground truth emotion's angle and its weighted average reported angle, and then averaging that value over the emotions within each quadrant. It is worth noting that only three emotions in the dataset and two emotions in the generated data had weighted average angles outside the correct quadrant. The final units were converted from degrees to units of emotion (20 emotions in 360°). The last columns, showing variance, were calculated by finding the weighted variance for each emotion (converted to units of emotion), and then averaging for each quadrant.

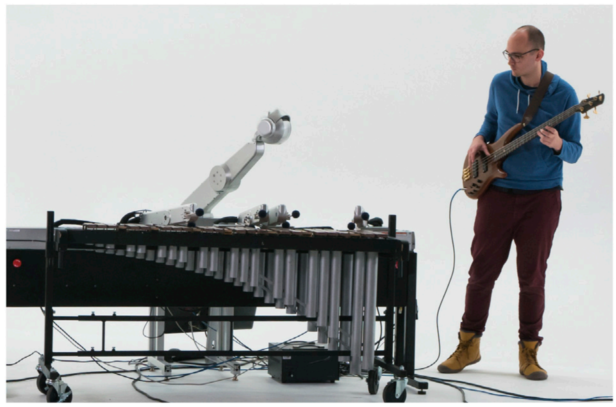


FIGURE 5 | Shimon the robotic marimba player.

Our results show that the generated phrases performed similarly to the dataset in terms of emotion classification. While the percentage of phrases identified in the correct quadrant is slightly lower for the generated phrases, the average difference and variance have similar values. Visually, the rose plots show that participants were able to largely identify the correct quadrant, having the most difficulty with Quadrant 4 (relaxing/serene) for both our collected dataset and generations.

5 EXPERIMENT

After creating the described prosody generation system we linked the system to our custom robotic platform Shimon. Shimon is a four-armed marimba playing robot that has been used for a wide range of musical tasks from improvisation (Hoffman and Weinberg, 2010) to film scores (Savery and Weinberg, 2018). **Figure 5** shows Shimon improvising with a human performer. To visually show Shimon voicing the prosody we copied a previous implementation used to link Shimon's gestures to human language for hip hop (Savery et al., 2020b).

For the experiment, we considered creativity through Boden's framework for computational creativity (Boden, 2009). Boden considers creativity as a balance between novelty and coherence, with expressivity playing a significant role in the process. This concept draws on the notion that a completely random idea could be considered novel, yet would lack coherence. Boden's framework was used to evaluate computational creativity in a number of previous works (Riedl and Young, 2010; Savery et al., 2020b).

We choose to compare musical prosody to a text-to-speech system for Shimon. Speech is very commonly used in robotics (Brooks et al., 2012; Niculescu et al., 2013) and is likely the primary form of audio interaction. Speech is often described as a way for replicating human to human communication (Crumpton and Bethel, 2016) and we believe would commonly be considered the default audio type for a robot such as Shimon.

Our experiment was designed to answer two research questions:

(1) Can emotional prosody improve a robot's creative output, as measured through novelty, coherence and expressivity when compared to a text-to-speech system?
(2) Can emotional prosody alter the perception of animacy, anthropomorphism, likeability and intelligence for a creative robot compared to a text-to-speech system?

For these research questions we developed two exploratory hypothesis, extending the work of Moore (2017), where voices matching the mode of interaction will improve the interaction. For research question 1 we hypothesize that when communicating using emotion-driven prosody, Shimon will achieve higher ratings for novelty, and expressivity with a significant result, while coherence will not have significant difference. We hypothesize this will occur since prosody will increase the image of Shimon as creative agent, but not alter coherence. This aligns with our design goals of addressing the habitability gap and aiming for a robot that interacts in a manner that matches its performance. For research question 2 we hypothesize that there will be no difference in perception of animacy, and anthropomorphism, however prosody will achieve a significant result for higher likeability. We believe that the extra functionality implied by a text-to-speech system will enhance the perceived intelligence.

5.1 Experimental Design

We conducted the experiment as a between-group study, with one group watching robotic interactions with a text-to-speech system and the other with our generative prosody system. The study was set up as an online experiment with participants watching videos of a musician interacting with Shimon. For the text-to-speech we used Google API with a US female voice (en-US-Wavenet-E) (Oord et al., 2016). We chose the voice model as it is easily implemented in real time and a widely used system.

The musical interactions involved six clips of a human improvising four measures, followed by Shimon responding with a four-measure-long improvisation. The improvisation was over a groove at 83 beats per minute, resulting in the improvisation lasting for about 23 s. Each improvisation was followed by a seven-second gesture and response from Shimon, either using text-to-speech or prosody. Both the speech and prosody used three high valence-low arousal and three low valence-low arousal phrases. The prosody or text-to-speech was overdubbed after recording allowing us to use identical musical improvisations from the human and robot. For text-to-speech we used phrases that designed by the author based on past interactions in rehearsal between human participants.

The high valence-low arousal text included the three phrases:

- Great work. What you played really inspired me to play differently. Could you hear how we were able to build off each others music?
- That was fun, it was good playing with you. I really liked hearing the music you played on keyboard, it worked well with what I played.
- Thanks so much for playing here with me, I thought what you played was really good. Let's keep playing together.

The low valence-low arousal text included the three phrases:

- Let's try it again soon, the more we play together the more we will improve. I'm going to listen to you really carefully next time
- That was a really good start, I enjoyed the way we interacted together. We should keep trying to work on it and get better.
- Did you listen to what I played? Do you think it worked well with what you played? The more we practice the better we can get.

Participants first completed a consent form outlining the process, and then read brief instructions on the experiment process. After watching three of the clips they were asked to rate them based on Boden's metrics, then repeated the process for the next three clips. Boden's metrics were rated on a seven point sliding scale. Participants were explicitly asked to rate the musical improvisation from the robot for each metric. Clips were randomly ordered for each participant. Additionally, a seventh clip was added as an attention check, which included an additional video. In this video, instead of sound, participants were asked to type a word that was asked for at the end of the survey.

After watching each interaction, participants rated animacy, anthropomorphism, likeability and perceived intelligence using the Godspeed measure (Bartneck et al., 2009). Each metric contained four or five sub-questions, which were averaged to

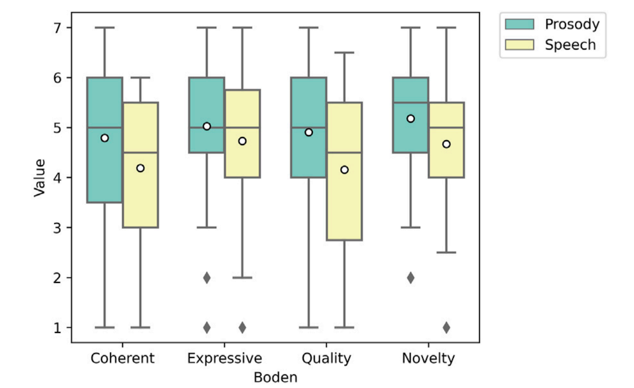


FIGURE 6 | Box plot of Boden's Creativity Metrics. Quality is significant, $p = 0.014$.

give an overall rating. To conclude the experiment, participants answered demographic questions and were given an open text response to comment on the robot or experiment.

We used Amazon Mechanical Turk (MTurk) to recruit participants who then completed the survey through Qualtrics. MTurk is a crowd-sourcing platform created by Amazon that allows individuals and businesses to hire users to complete surveys. Participants were paid $2.00 upon completion of the survey, which took around 10 min. We allowed only MTurk Masters to participate, and required a successful job rate of 90%. We also monitored time to complete overall, and time spent to complete each question. We recruited 106 initial participants, four of whom failed the attention check. An additional two participants were disqualified as they completed the survey in under 5 min. As participants failed the attention check a new spot was immediately opened allowing us to reach 100 participants. In total we included data from 50 participants who heard the text-to-speech system and 50 who heard the prosody system. The mean age of participants was 44, ranging from 25 to 72, with a standard deviation of 11. The majority of participants were based in the United States (89) with the remaining in India (11). We found no difference in comparisons of the results between each country. Considering the gender of each participant, 39 identified as female, 60 as male and one as non-binary.

5.2 Results

Our analysis was conducted with a Jupyter Notebook, running directly on the exported CSV from qualtrics. Libraries for analysis included NumPy, and SciPy.stats.

5.2.1 Creativity

Prosody had a higher mean for coherence 4.80 (*std* = 1.31), novelty 5.18 (*std* = 1.30), and quality 4.95, (*std* = 1.68) compared to speech with the means 4.19 (*std* = 1.56), 4.64 (*std* = 1.24), and 4.14 (*std* = 1.37). Prosody had effect sizes of 0.40 for coherence, 0.43 for novelty, and 0.56 for quality indicating a medium size effect calculated using Cohen's D. For expressivity prosody had an effect size of 0.25, indicating a small effect size. After conducting a pairwise *t*-test across categories were significant with the results, coherence ($p = 0.041$), novelty ($p = 0.040$), and quality ($p = 0.014$). After a Bonferroni-Holm correction for multiple comparisons, only quality remained significant with ($p = 0.014$) while coherence ($p = 0.12$) and novelty ($p = 0.12$) where no longer significant. For expressivity, prosody only had a slightly higher mean which was not significant ($p > 0.05$). **Figure 6** shows a box plot of all Boden's metrics.

5.2.2 Godspeed

For the Godspeed metrics we first calculated Cronbach's alpha for each question subset. This resulted in animacy (0.86), anthropomorphism (0.88), likeability (0.92), perceived intelligence (0.89). This shows high internal reliability across all metrics. Prosody had an effect size for each metric as animacy (0.16), anthropomorphism (0.08), likeability (0.85) and perceived intelligence (0.54), measured with Cohen's D.

Prosody had a slightly higher mean for animacy 3.56 (*std* = 0.88) compared to speech 3.44 (*std* = 0.75). Prosody also had a slightly higher rating for anthropomorphism 3.14 (*std* = 0.99), compared to speech 3.08 (*std* = 0.885). After running a pairwise t-test neither animacy or anthropomorphism were significant. Prosody had a higher mean for likeability, 4.38 (*std* = 0.89) compared to 3.94 (*std* = 0.52) and showed a significant result ($p = 0.002$) in a pairwise *t*-test, which remained significant after a Bonferroni-Holm correction for multiple comparison ($p = 0.011$). For perceived intelligence, prosody 4.10 (*std* = 0.82) outperformed speech 3.72 (*std* = 0.70), with a significant result ($p = 0.014$) which remained significant after correction ($p = 0.042$). **Figure 7** shows a box plot of all Godspeed metrics.

6 DISCUSSION AND FUTURE WORK

6.1 Research Question 1

Overall, our results indicated that the communication method outside of performance made a significant difference in participant ratings of creativity. The higher ratings for novelty and quality supported our hypothesis that prosody would outperform speech, however we did not expect coherence to improve with prosody as well. Surprisingly, we found no

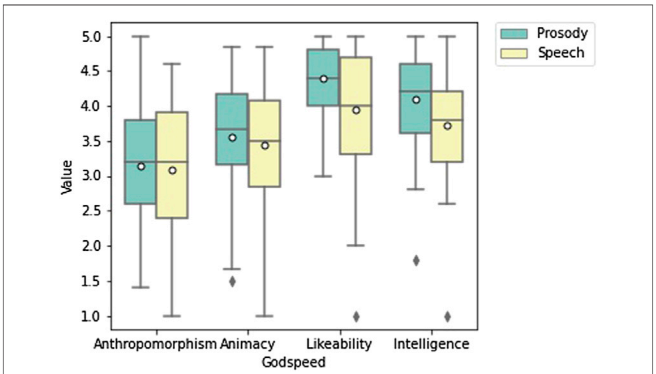


FIGURE 7 | Box plot of Godspeed Metrics. Likeability and Intelligence are significant $p = 0.011$ and $p = 0.042$.

significant difference between voice type for expressivity and additionally expressivity only had a small effect size. This did not support our hypothesis as we had expected prosody to create the impression of a more expressive robot.

Further research is required to understand why the perception expressivity, as a creativity trait, did not change based on the voice used. One possible reason is that participants believed a robot that could use language was capable of a wide range of expression, much like the addition of prosody. Alternatively, expressivity is a feature that is not easily altered by the form of interaction post-performance.

The relation between each creativity rating cannot be easily simplified, and there is no correct answer to what rating a performance should receive for coherence or novelty. We expected that the prosody system would receive higher ratings for novelty, but not coherence. We believe that the higher ratings for coherence may have come from the system acting as a unified robot, with its communication functioning in the same manner as its performance.

6.2 Research Question 2

Our results for likeability matched our hypothesis that prosody would outperform speech. Perceived intelligence ratings however did not support our hypothesis as we had predicted language would be interpreted as having a higher intelligence. It was reasonable to assume that with text-to-speech and the ability to speak a language, Shimon would have been perceived as more intelligent. We found that the system with prosody was considered more intelligent, despite not communicating linguistically. This can be explained by the assumptions that moving towards the habitability gap will create a disjointed perception of the robot. A possible conclusion was that participants understood there was not a deep knowledge of language, whereas musical phrases implies a deeper musical intelligence.

6.3 Text Responses

We found no distinct variation in text responses between the speech and prosody group. Overall 92 participants chose to respond, with responses ranging from one sentence to four sentences. From the speech group only one participant mentioned the voice, writing “I enjoyed the robot, especially when she spoke to the pianist” (gender added by participant). In the prosody responses four participants mentioned the voice, but only in passing, such as the voice was “cute.” The vast majority of response rated the musical responses and generations, with the majority positive such as “I liked the robot and I like the robots music more than the humans,” and “Nice to listen to.” The negative comments tended to focus on the inability of robots in general to play music or be creative such as “It could play notes, but it lacked creativity.”

6.4 Generative Process

Our dataset used interpretation of emotions from one vocalist. While this had the benefit of consistency throughout phrases, in future work we intend to gather data from a larger number of musicians and to evaluate how well the model can generalize. We also plan to have other robots communicating through prosody using data from different vocalists.

We plan to further investigate timbre and its potential application to the generation process. We also intend to study which features of the phrases influenced participants’ choice in selecting an emotion. For example, exploring whether there is a difference in emotion classification accuracy for the melody of the generated phrases alone, in comparison with emotionally-sampled audio as we used here. Future work will also include more extensive studies using the generated prosody in human-robot interactions. This will take place between varying group sizes from one human and robot, to groups of humans and robots with different embedded personalities. We expect for emotional musical prosody to enable many future collaborations between human and robot. Our overall accuracy presented in **Table 1** shows consistent results in the mid 50%. We believe this accuracy is acceptable for our current system, as the average variance and average difference are both close to two across all categories, implying that the primary errors in identification where small, such as mistaking love for admiration. For our experiment in particular we only used two quadrants, and were also able to choose only specific emotions that scored over 80% accuracy.

In both the original dataset and generated material participants had the lowest accuracy identifying the fourth quadrant emotions. Our results are not easily compared to other generative systems as the fourth quadrant emotions are rarely used in robotic studies Savery and Weinberg (2020). This is partly because common classification systems such as Ekman’s discrete classes do not include anything in the fourth quadrant. We also believe these emotions tend to be less easily displayed externally as they are low arousal and closer to neutral emotions. In future work we aim to consider methods to better develop the fourth quadrant emotions.

6.5 Limitations

We compared one text-to-speech system with one musical prosody system on one robotic platform. In future work we aim to compare further audio systems, to expand understandings of why different metrics showed significant results. It is possible that varying the speech used would alter the final ratings. Nevertheless, we believe that the range of metrics that did prove significant show that this is an important first step in understanding how communication between core creative tasks can shape the perception of a robot.

We were only able to compare two forms of communication in a the constrained scenario consisting of directly after a musical interaction. To restrict our experiment to two groups we did not compare prosody to moments where the robot did not interact at all. We believe that by its nature a robot such as Shimon is always interacting and its presence can alter humans actions (Hoffman et al., 2015), leading us to believe that no movement or audio is its own form of interaction. In future research we intend to analyze the impact of musical prosody compared to no interaction in a longer performance.

This study was conducted online through video, which comes with benefits and drawbacks. As we were running online we were able to gather many more participants than would have been possible in person. Similar HRI studies have shown no difference in online replication of certain studies (Woods et al., 2006), and we believe our method was constrained to a point that would be replicated in an in-person study. We did not include a manipulation check in our study, however our analysis of the text responses indicated that participants did not identify the independent variable between groups.

The range of participants included in the study also adds some limitations. Our primary goal was to understand how changes to a creative system would generalize across a broad population. We did not factor in concerns between cultural groups that may take place, such as between Japan and United States (Fraune et al., 2015), however our study did not find any significant variation between origin country. Additionally, our ability to generalize is restricted by only collecting participants on MTurk, who it has been shown do not always represent standard population samples, such as in the case of participants health status (Walters et al., 2018). Finally, our sample size of 106 participants was under the total that would be required to detect an effect size of 0.50 with 0.80 power at an alpha level of 0.05, which requires a sample size of 128.

7 CONCLUSION

The paper presents a new generative system for emotional musical prosody that is implemented in Shimon, a creative robot. We explore how a robot's response outside of its key creative task—such as musical improvisation—alters the perception of the robot's creativity, animacy, anthropomorphism, perceived intelligence, and likeability. Our research questions focus on how prosody compares to text-to-speech in a creative system for each of these HRI metrics.

We found that by addressing the habitability gap we were able to increase user ratings for the key creativity ratings; novelty and coherence, while maintaining ratings for expressivity across each implementation. Our results also indicated that by communicating in a form that relates to the robot's core functionality, we can raise likeability and perceived intelligence, while not altering animacy or anthropomorphism. These findings clearly indicate the impact of developing interactions surrounding a creative performance, such as initial meetings and gaps between creative interaction.

Our results present wide ranging implications and future concepts for the development of creative robots. The importance of design outside primary tasks should not only be considered for creative robots, but across HRI. These findings indicate that embodiment and external design choices alter not only the impression of the robot, but the impression of its primary functions. We also believe this work indicates the importance of audio design, and the impact on perception that changes to audio alone can have on a system. By designing audio for the system task and not relying on default audio methods it is possible to drastically change the perception of a robotic system.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Georgia Tech Institutional Review Board. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

RS developed the concept and paper with input from all authors, and designed and conducted the experiment. LZ and RS collected the dataset and designed the generative system. GW oversaw the design, implementation, and writing of the described research.

FUNDING

This material is based upon work supported by the National Science Foundation under Grant No. 1925178.

ACKNOWLEDGMENTS

Thanks to Heather Song and Amit Rogel who assisted with creation of the gesture stimuli and interaction for the experiment.

REFERENCES

Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *Int. J. Soc. Robotics* 1, 71–81. doi:10.1007/s12369-008-0001-3

Bates, J. (1994). The Role of Emotion in Believable Agents. *Commun. ACM* 37, 122–125. doi:10.1145/176789.176803

Boden, M. A. (2009). Computer Models of Creativity. *AIMag* 30, 23. doi:10.1609/aimag.v30i3.2254

Briot, J.-P., Hadjeres, G., and Pachet, F. D. (2017). Deep Learning Techniques for Music Generation–A Survey. Available at: http://arxiv.org/abs/1709.01620.

Brooks, D. J., Lignos, C., Finucane, C., Medvedev, M. S., Perera, I., Raman, V., et al. (2012). "Make it So: Continuous, Flexible Natural Language Interaction with an Autonomous Robot". in proceedings Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence. doi:10.1145/2157689.2157827

Brunner, G., Konrad, A., Wang, Y., and Wattenhofer, R. (2018). Midi-vae: Modeling Dynamics and Instrumentation of Music With Applications to Style Transfer. Available at: http://arxiv.org/abs/1809.07600. doi:10.1109/ictai.2018.00123

Coyne, A. K., Murtagh, A., and McGinn, C. (2020). "Using the Geneva Emotion Wheel to Measure Perceived Affect in Human-Robot Interaction," in Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction—HRI'20. (New York, NY, United States: Association for Computing Machinery), 491–498. doi:10.1145/3319502.3374834

Crumpton, J., and Bethel, C. L. (2016). A Survey of Using Vocal Prosody to Convey Emotion in Robot Speech. *Int. J. Soc. Robotics* 8, 271–285. doi:10.1007/s12369-015-0329-4

Dautenhahn, K., Walters, M., Woods, S., Koay, K. L., Nehaniv, C. L., Sisbot, A., et al. (2006). "How May I Serve You? A Robot Companion Approaching a Seated Person in a Helping Context," in Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction—HRI '06 (New York, NY, United States: Association for Computing Machinery), 172–179. doi:10.1145/1121241.1121272

Ferreira, L., and Whitehead, J. (2019). Learning to Generate Music with Sentiment. *ISMIR*, 384–390.

Fischer, K. (2019). Why Collaborative Robots Must be Social and Even Emotional Actors. *Techné: Res. Philos. Technol.* 23, 270–289. doi:10.5840/techne20191120104

Fraune, M. R., Kawakami, S., Sabanovic, S., De Silva, P. R. S., and Okada, M. (2015). "Three's Company, or a Crowd?: The Effects of Robot Number and Behavior on Hri in japan and the USA," in Conference: International Conference on Robotics Science and System (RSS2015), Rome, Italy. doi:10.15607/rss.2015.xi.033

Fukuda, H., Shiomi, M., Nakagawa, K., and Ueda, K. (2012). "midas Touch" in Human-Robot Interaction: Evidence from Event-Related Potentials during the Ultimatum Game," in Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction. 131–132.

Gleeson, B., MacLean, K., Haddadi, A., Croft, E., and Alcazar, J. (2013). "Gestures for Industry Intuitive Human-Robot Communication from Human Observation," in Proceedings 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI): (IEEE), 349–356.

Goetz, J., Kiesler, S., and Powers, A. (2003). "Matching Robot Appearance and Behavior to Tasks to Improve Human-Robot Cooperation," in Proceedings The 12th IEEE International Workshop on Robot and Human Interactive Communication: (IEEE), 55–60.

Hoffman, G.., Forlizzi, J., Ayal, S., Steinfeld, A., Antanitis, J., Hochman, G., et al. (2015). "Robot Presence and Human Honesty: Experimental Evidence," in Proceedings 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI): (IEEE), 181–188.

Hoffman, G., and Weinberg, G. (2010). "Synchronization in Human-Robot Musicianship," in Proceedings 19th International Symposium in Robot and Human Interactive Communication, 718–724. doi:10.1109/ROMAN.2010.5598690

Huang, C. Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., et al. (2018). Music Transformer. Available at: http://arxiv.org/abs/1809.04281.

Jones, R. (2013). Communication in the Real World: An Introduction to Communication Studies. Twin Cities, MN: The Saylor Foundation.

Kim, J. W., Salamon, J., Li, P., and Bello, J. P. (2018). "Crepe: A Convolutional Representation for Pitch Estimation," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP): (IEEE), 161–165.

Kingma, D. P., and Welling, M. (2013). Auto-encoding Variational Bayes. Available at: http://arxiv.org/abs/1312.6114.

Mavridis, N. (2015). A Review of Verbal and Non-verbal Human-Robot Interactive Communication. *Robotics Autonomous Syst.* 63, 22–35. doi:10.1016/j.robot.2014.09.031

Moon, A., Parker, C. A., Croft, E. A., and Van der Loos, H. M. (2011). "Did You See it Hesitate?-Empirically Grounded Design of Hesitation Trajectories for Collaborative Robots," in IEEE/RSJ International Conference on Intelligent Robots and Systems: (IEEE), 1994–1999.

Moore, R. K. (2017). "Is Spoken Language All-Or-Nothing? Implications for Future Speech-Based Human-Machine Interaction," in Dialogues with Social Robots: (Springer), 281–291. doi:10.1007/978-981-10-2585-3_22

Niculescu, A., van Dijk, B., Nijholt, A., Li, H., and See, S. L. (2013). Making Social Robots More Attractive: the Effects of Voice Pitch, Humor and Empathy. *Int. J. Soc. Robotics* 5, 171–191. doi:10.1007/s12369-012-0171-x

Oord, A. V. D., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., et al. (2016). Wavenet: A Generative Model for Raw Audio. Available at: http://arxiv.org/abs/1609.03499.

Pires, J. N., and Azar, A. S. (2018). Advances in Robotics for Additive/hybrid Manufacturing: Robot Control, Speech Interface and Path Planning. *Ind. Robot: Int. J.*

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive Development, and Psychopathology. *Development psychopathology* 17, 715. doi:10.1017/s0954579405050340

Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). Stochastic Backpropagation and Approximate Inference in Deep Generative Models. Available at: http://arxiv.org/abs/1401.4082.

Riedl, M. O., and Young, R. M. (2010). Narrative Planning: Balancing Plot and Character. *jair* 39, 217–268. doi:10.1613/jair.2989

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., and Eck, D. (2018). A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music. Available at: http://arxiv.org/abs/1803.05428.

Sacharin, V., Schlegel, K., and Scherer, K. (2012). Geneva Emotion Wheel Rating Study (Report). Geneva, Switzerland: University of geneva. Swiss Center for Affective Sciences.

Saunderson, S., and Nejat, G. (2019). How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human-Robot Interaction. *Int. J. Soc. Robotics* 11, 575–608. doi:10.1007/s12369-019-00523-0

Savery, R., Rose, R., and Weinberg, G. (2019a). "Establishing Human-Robot Trust through Music-Driven Robotic Emotion Prosody and Gesture", in 2019 28th IEEE International Conference on Robot and Human Interactive Communication: (RO-MAN IEEE), 1–7.

Savery, R., Rose, R., and Weinberg, G. (2019b). "Finding Shimi's Voice: Fostering Human-Robot Communication with Music and a Nvidia Jetson Tx2," in Proceedings of the 17th Linux Audio Conference.

Savery, R., and Weinberg, G. (2020). "A Survey of Robotics and Emotion: Classifications and Models of Emotional Interaction," in 2020 29th IEEE International Conference on Robot and Human Interactive Communication: (RO-MAN) (IEEE), 986–993.

Savery, R., and Weinberg, G. (2018). Shimon the Robot Film Composer and Deepscore. *Proc. Comp. Simulation Musical Creativity* 5.

Savery, R., Zahray, L., and Weinberg, G. (2020a). "Emotional Musical Prosody for the Enhancement of Trust in Robotic Arm Communication," in Trust, Acceptance and Social Cues in Human-Robot Interaction: 29th IEEE International Conference on Robot & Human Interactive Communication. doi:10.1109/ro-man47096.2020.9223536

Savery, R., Zahray, L., and Weinberg, G. (2020b). "Shimon the Rapper: A Real-Time System for Human-Robot Interactive Rap Battles," International Conference on Computational Creativity. doi:10.1109/ro-man47096.2020.9223536

Scherer, K. R., Trznadel, S., Fantini, B., and Sundberg, J. (2017). Recognizing Emotions in the Singing Voice. *Psychomusicology: Music, Mind, and Brain* 27, 244. doi:10.1037/pmu0000193

Sohn, K., Lee, H., and Yan, X. (2015). Learning Structured Output Representation Using Deep Conditional Generative Models. *Adv. Neural Inf. Process. Syst.*, 3483–3491.

Sturm, B. L., Santos, J. F., Ben-Tal, O., and Korshunova, I. (2016). Music Transcription Modelling and Composition Using Deep Learning. Available at: http://arxiv.org/abs/1604.08723.

Walters, K., Christakis, D. A., and Wright, D. R. (2018). Are Mechanical Turk Worker Samples Representative of Health Status and Health Behaviors in the U.S.?. *PloS one* 13, e0198835. doi:10.1371/journal.pone.0198835

Walters, M. L., Dautenhahn, K., Te Boekhorst, R., Koay, K. L., Kaouri, C., Woods, S., et al. (2005). "The Influence of Subjects Personality Traits on Personal Spatial Zones in a Human-Robot Interaction Experiment," in ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication: (IEEE), 347–352.

Woods, S., Walters, M., Koay, K. L., and Dautenhahn, K. (2006). "Comparing Human Robot Interaction Scenarios Using Live and Video Based Methods: Towards a Novel Methodological Approach," in 9th IEEE International Workshop on Advanced Motion Control: (IEEE), 750–755.

Wu, J., Hu, C., Wang, Y., Hu, X., and Zhu, J. (2020). A Hierarchical Recurrent Neural Network for Symbolic Melody Generation. *IEEE Trans. Cybern* 50, 2749–2757. doi:10.1109/TCYB.2019.2953194

Yang, L. C., Chou, S. Y., and Yang, Y. H. (2017). Midinet: A Convolutional Generative Adversarial Network for Symbolic-Domain Music Generation. Available at: http://arxiv.org/abs/1703.10847.

Zhao, K., Li, S., Cai, J., Wang, H., and Wang, J. (2019). "An Emotional Symbolic Music Generation System Based on Lstm Networks," in 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). 2039–2043.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

HYPOTHESIS AND THEORY
published: 30 April 2021
doi: 10.3389/frobt.2021.662182

Creative Action at a Distance: A Conceptual Framework for Embodied Performance With Robotic Actors

Philipp Wicke and Tony Veale*

School of Computer Science, University College Dublin, Belfield, Ireland

Acting, stand-up and dancing are creative, embodied performances that nonetheless follow a script. Unless experimental or improvised, the performers draw their movements from much the same stock of embodied schemas. A slavish following of the script leaves no room for creativity, but active interpretation of the script does. It is the choices one makes, of words and actions, that make a performance creative. In this theory and hypothesis article, we present a framework for performance and interpretation within robotic storytelling. The performance framework is built upon movement theory, and defines a taxonomy of basic schematic movements and the most important gesture types. For the interpretation framework, we hypothesise that emotionally-grounded choices can inform acts of metaphor and blending, to elevate a scripted performance into a creative one. Theory and hypothesis are each grounded in empirical research, and aim to provide resources for other robotic studies of the creative use of movement and gestures.

Keywords: robotics, computational creativity, embodiment, storytelling, spatial movement

OPEN ACCESS

Edited by:
Amy LaViers, University of Illinois at Urbana-Champaign, United States

Reviewed by:
Safinah Ali, Massachusetts Institute of Technology, United States
Alexandra Baybutt, Buckinghamshire New University, United Kingdom

****Correspondence:***
Philipp Wicke
philipp.wicke@ucdconnect.ie

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *31 January 2021*
Accepted: *12 April 2021*
Published: *30 April 2021*

Citation:
Wicke P and Veale T (2021) Creative Action at a Distance: A Conceptual Framework for Embodied Performance With Robotic Actors. Front. Robot. AI 8:662182. doi: 10.3389/frobt.2021.662182

1 INTRODUCTION

Embodied performances on a stage often start with a script. Performers can slavishly follow this script, like a computer executing a computer program, or they can interpret its directives as they see fit. Only by doing the latter can a performer be said to deliver a "creative" performance.

A performance is a conceptual scheme turned into physical action. When concepts become movements, movements suggest meanings and meanings evoke concepts in the minds of an audience. Since every link in this chain is under-determined, creativity can insinuate itself into every part of the meaning-making process. The physical actions of a performance suggest meanings, or reinforce what is also communicated with words, so the most effective actions tap into an audience's sense of familiarity, obviousness and conceptual metaphor. In this paper, we consider story-telling as a performance that combines linguistic (spoken) and physical (embodied) actions. Our embodied actors can communicate a tale by narrating it, or by acting it out, or as an ensemble of agents that do both. Our focus is unique for a number of reasons. First, our embodied actors are robots, not humans, although they aim to move, pose and gesture much as humans do. Second, the tales they tell are not spun by a human, but generated by a machine in an act of computational creativity. An AI system that controls the writing process, the telling process and the acting process can thus be used to explore the ties between concepts, words and embodied actions in a creative, performative setting. Third, our robots can interpret the written script just as actors interpret a film script. They can literally depict the actions through movement and gesture, or they can interpret the actions of the script metaphorically. This flexible reading of the script allows metaphor to shape its embodied interpretation, fostering creativity in the physical enactment of the story. In short, we explore here how interpretation is infused with emotionally-grounded choice to appreciate and to achieve embodied creativity in a system for performing machine-generated stories.

Story-telling is just one kind of embodied performance. We humans use our bodies to tell jokes, engage in animated conversation, and communicate feelings in play and in dance. Starting from a narrative perspective, with a system designed to support the performance of computer-generated stories with computer-controlled robotic actors, we set out to generalize our approach and create a framework for embodied communication that can support multiple types of performance. Key to this approach is the meaning-making potential of physical acts, which we ground in image-schematic models of language. As we will show here, story-telling provides an ideal basis for empirically testing our hypotheses, but our aim is to broaden the framework to accommodate new possibilities and new kinds of performance.

We adopt a bottom-up approach to unifying theory and practice, in which an implemented AI system supports the empirical studies that motivate our hypotheses, before we generalize those hypotheses into a combinatorial framework for embodied meaning-making. We begin by surveying the state-of-the-art in robotic performance to define a taxonomy that accommodates humanoid movements from walking to posing and gesturing. Although physical actions are not words, deliberate physical actions do have a semiotic component that we will analyze here. So, by exploring robotic enactment in a storytelling context, we can identify the semantic units of movement and their cognitive-linguistic underpinnings in image schemas and conceptual metaphors. Ultimately, our goal is to identify the points of contact between action and meaning where creativity – and in particular, machine creativity – can blossom.

In our *Performance Framework* (**Section 3**), we outline which movements can be executed in parallel or in series, to convey meanings of their own or to augment the meaning of the spoken dialogue. In addition, we will consider the properties of physical actions to identify those that are additive (when compounded movements achieve a cumulative effect), persistent (when a movement has a lasting effect on the physical relationship between actors), and summative (allowing an action to summarize what has already occurred). For example, the meaning conveyed by one actor stepping away from another intensifies with each additional step. The action and its meaning are also persistent, since, unlike gestures, stepping away does not necessitate a subsequent retraction of the action. After multiple steps, the resulting distance between actors (and characters) is the sum of all steps, and so conveys a global perspective.

We distinguish between locomotive, spatial movements (hereinafter spatial movement) along a stage, postural reorganizations of the body, and gestures made with the hands, arms and upper-body to communicate specific intents. For gestures, we also discriminate pantomimic or iconic gestures (which play-act a meaning, e.g., *using an invisible steering wheel to signify driving*) from more arbitrary actions (which may use metonymy to depict culturally-specific actions, such as *bending the knee to propose*), from those that instantiate a conceptual metaphor to achieve their communicative intent. The framework formally integrates each of these forms of physical meaning-making, and constrains how they work with each other in the realization of a coherent performance.

Connecting the underlying script with the performance requires an appropriate choice of the movements to be enacted. When a script dictates the actions, there is no space for choice. Likewise, when a script provides simple disjunctive choices – do *this* or *that* – it allows a performer to explore the space of possible stories without regard for the emotions of the characters. The actor's capacity to interpret the script should consider these emotions and how they can shape the performance. By creating choices at the time of performance, an interpretation can look at the unfolding narrative so far and shape the course that the actor will take. Since these choices are guided by the character's emotional valence at any given moment, we introduce an emotional layer between the level of scripted actions (what happens next) and the expressive level (physical movements and spoken words). This new layer annotates the emotions implied for each action and movement, to inform the actors about the emotional resonances of their choices. By considering the influence of earlier actions in the plot, choices can be made in the moment, to reflect an interpretation of how characters should be feeling and acting. The *Interpretation Framework* (**Section 4**) provides the tools to a performer to make deliberate use of gesture and space for an emotionally-informed performance.

Since the AI system automatically maps tales from pre-generated texts onto physical performances, we can use these performances as the basis of empirical studies that explore whether audiences intuitively appreciate the deliberate use of space and gesture in a performance. More specifically, we look at whether coherent usage is as appreciated as incoherent use, and whether the schematic use of space in a cumulative, summative fashion is as appreciated by audiences as the use of transient, culturally-specific gestures. We interpret the results of those studies with respect to the frameworks presented here. Both studies (described in **Section 5**) have been conducted by recording robotic performances under the coherent and incoherent conditions, and then eliciting crowd-sourced ratings of those performances. Each participant is shown short videos of plot segments that feature relevant movements, and each is asked to rate the performance on a customized HRI questionnaire. As such, we intend to contribute to multiple areas of interdisciplinary research with this framework: not just automated storytelling (as built on automated story-generation) and embodied performance using robots, but the study of expressive gestures and physical meaning-making more generally, across a diversity of settings. While we evaluate a rather specific use of the framework in the story domain, we will provide a taxonomy and a terminology that will foster further interdisciplinary research in the areas that contribute to it.

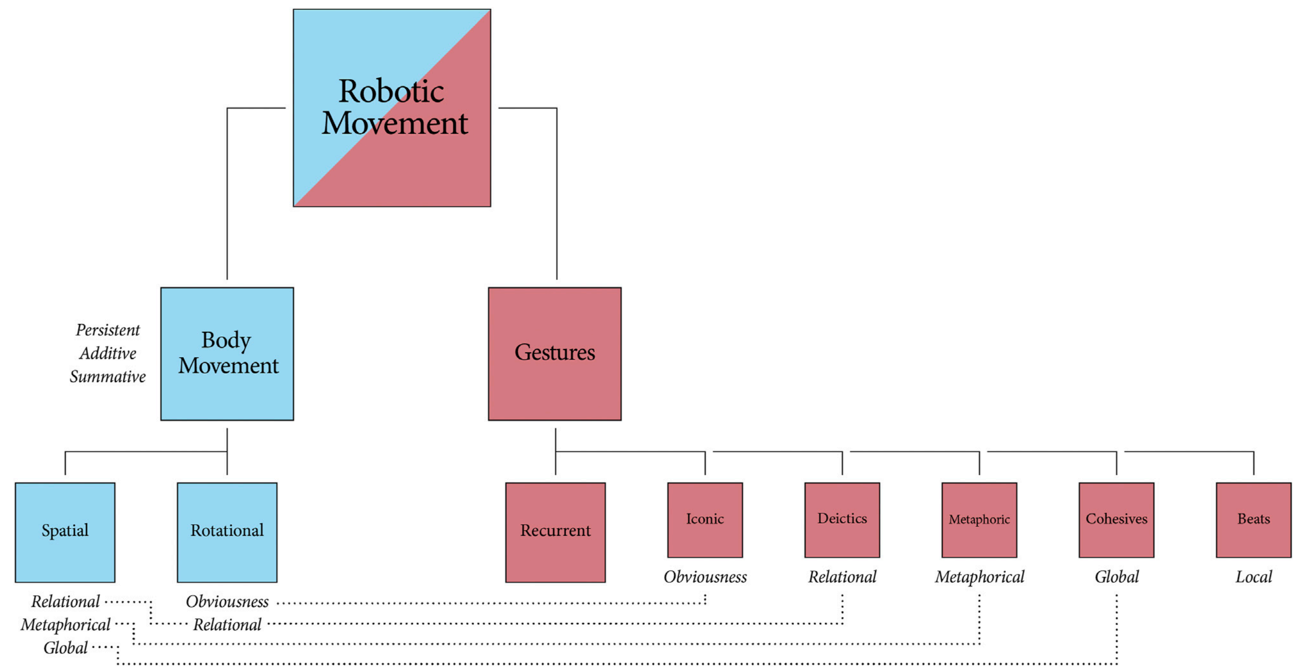


FIGURE 1 | A taxonomy of robotic movements and their properties. Nodes to the left and right depict broad types of robotic movement and their specific sub-types (rotational, spatial, iconic gestures, narrative beats, etc). Shared properties are indicated with horizontal connections (dashed lines).

2 BACKGROUND AND RELATED WORK

Although this framework is evaluated in a storytelling setting, it is applicable in different contexts of choreographed robotic interaction. Thus, we apply a new interpretation to existing data and argue that the framework has interdisciplinary relevance to other researchers in robotics.

Robots can make use of a variety of different modalities, each of which has been studied in different contexts: gaze (Mutlu et al., 2012; Andrist et al., 2014), facial expression (Reyes et al., 2019; Ritschel et al., 2019), voice (Niculescu et al., 2013), gesture (Pelachaud et al., 2010; Ham et al., 2011) and movement (Shamsuddin et al., 2011). The focus of our framework is on the movement of the robot, including both gestures and spatial movements (walking to and fro). Gestures have been extensively studied in linguistics and human-robot interaction, while spatial movement that concerns the whole body has been studied with social robots for comedy (Katevas et al., 2014), theater/improvisation (Bruce et al., 2000; Knight, 2011), and dance (LaViers and Egerstedt, 2012; Seo et al., 2013). Our consideration of related work thus takes an interdisciplinary look at various definitions and properties of gestures and holistic body movements, and derives a basis for characterizing the properties of spatial movements and gestures with reference to a series of empirical studies. We conclude by outlining the advantages of this framework for robotic performance.

2.1 Robotic Movement

We distinguish two classes of performative movement: local movement or gesture, typically with the arms, hands or head, and spatial relocation of the whole body. Gestures can arguably involve the whole body (as in e.g., air marshalling), while bodily locomotion can involve gestures while in motion. Yet the literature on gestures mostly confines gestures to the upper torso, arms and head (McNeill, 2008), while McNeill's widely used reference frame for gesture space depicts a sitting human with only the upper torso, arms and head in play (McNeill, 1992). Restriction to the upper body implies a locality of movement, while shifting the body in space has proximity effects that are global and relational. Locality and relativity are just two of multiple properties that reinforce a distinction between locomotive movements and gestures.

The taxonomic diagram in **Figure 1** defines the specific gestures and body movements we consider for our taxonomy. The top-most generalization *Robotic Movement* is split into *Body Movement* and *Gestures*, and each sub-type is linked to even more specific sub-types (vertically) and properties (horizontally). Although presented top-down, the taxonomy is a bottom-up approach that builds on a schematic basis for movement types to derive combinations and well-defined tools for roboticists and gesture researchers. Some related studies have looked at human motion in order to model movement dynamics (Bregler, 1997; Del Vecchio et al., 2003), while others have utilized computational models to simulate movement styles (Brand and Hertzmann, 2000; LaViers and Egerstedt, 2012).

2.1.1 Gestures

In addition to considering the spatial trajectories of gestures, we must also look at their expressive role in communicating meaning. Empirical work by McNeill (1985), Bergen et al. (2003) and Hauk et al. (2004) has shown that gestures are an important instrument of human communication. It has also been argued that gestures are always embedded in a social, ideological and cultural context, and as such, they infuse our conversations with a contextual semantics (Bucholtz and Hall, 2016). Although some researchers have proposed a unified methodology for the semantic study of gestures (Mittelberg, 2007), there is as yet no clear consensus around a single framework. Studies that focus on the timely execution of gestures, such as those exploring gesture recognition (Kettebekov and Sharma, 2001; Sharma et al., 2008), follow Kendon's approach to the separation of gestures into *preparation*, *stroke* and *retraction* phases (Kendon, 1980). Here, we note that the necessary *retraction* after a gesture makes the gesture transient and ephemeral, so that the posture of the performer is the same before and after the gesture is performed. A broader classification which has been adopted in many studies is provided by McNeill (1992):

- **Iconic:** A gesture resembles what is denotes. Example: Shadow boxing when talking about a fight.
- **Deictics:** A pointing gesture may refer to another object. Example: Pointing at another actor on stage.
- **Metaphoric:** A figurative gesture should not be taken literally, yet it communicates a truth about the situation. Example: Showing a trajectory with the hand when talking about a trip.
- **Cohesives:** A cohesive gesture binds two temporally distant but related parts of a narrative. Example: Making the same hand movement whenever the same character appears.
- **Beats:** A gesture marks narrative time. Example: A rhythmic arm movement indicates time passing.

The class of *Iconic* gestures requires that users recognize the iconicity of a gesture when it is performed by a robot. A study by Bremner and Leonards (2016) shows that iconic gestures performed by a robot can be understood by humans almost as well as those performed by humans. Another study, conducted by Salem et al. (2011), suggests that human evaluation of a robot is more positive when it uses iconic, referential and spatial gestures in addition to speech. Regarding spatial and referential gestures, it has been argued that gestures are primarily used to augment non-visuospatial speech communication with visuospatial information (McNeill, 1992). In the five classes of gesture above, most can convey some visuospatial information, but *Deictic* gestures do so by definition. *Deictics* play a crucial role in human to human communication by supporting direct reference to visual and non-visual objects (Norris, 2011). It has also been shown that robotic deictic gestures can shift our attention in much the same way as human uses of these gestures (Brooks and Breazeal, 2006). The level of abstraction in *Metaphoric* gestures is generally higher than that of *Iconic* and *Deictic* gestures, and there is evidence to suggest that distinct integration processes apply to these different classes of gesture in the human brain (Straube et al., 2011). Metaphors exploit familiar source domains, so the same gestural movement can be metaphorical in one speech context and iconic in another. For example, the gesture "*raising one arm above the head with a horizontal, open hand*" is iconic when it accompanies the sentence "*The plane flew way above the clouds*", and metaphorical when it accompanies "*She is way out of your league*". A study by Huang and Mutlu (2013) investigated four of McNeill's gesture classes (all but *Cohesives*) as used by interacting humanoid robots in a narrative context. Those authors evaluate each gesture type on several fronts: information recall, perceived performance, affective evaluation, and narration behavior. In their study, *Deictics* are shown to improve information recall relative to other gestures, while *Beats* lead to improvements in effectiveness.

There are observable overlaps between the reference framework used within spoken language and the reference framework used with gestures (Cienki, 2013a). For example, if an event occurs to the left of a person, that person is more inclined to gesture to their left when retelling the event (McNeill, 1992). While this appears to hold for most Indo-European languages, there are some culturally dependencies. Speakers of the Mayan language Tzeltal use an absolute spatial reference framework for both speech and gesture, so if an event occurs to the west of a Tzeltal speaker, they are inclined to point west when they later tell of it (Levinson, 2003). Another example of cultural diversity is found in the Aymaran language. The Aymara people of the Bolivian Andes refer to future events by pointing *behind* rather than *ahead* of themselves (Núñez and Sweetser, 2006). However, some gestures appear relatively stable across cultures when there is a consistent, well-established link from form to meaning (Ladewig, 2014). These recurrent gestures often serve a performative role, and fulfill a pragmatic function when they work on the level of speech (Müller et al., 2013). We exclude this class of gestures from our movement framework, since we focus here on gestural meaning-making that is parallel to, and not so easily tangled up with, speech. Nonetheless, for the sake of completeness, the recurrent gestures are depicted atop the other gesture classes in **Figure 1**.

A variety of studies have looked at gestures in human-robot interaction. Ham et al. (2011) evaluated a storytelling robot with a set of 21 handcrafted gestures and 8 gazing behaviors. Csapo et al. (2012) presented a multi-modal Q&A-dialog system for which they implemented 6 discourse-level gestures, much as Häring et al. (2011) had earlier presented a multi-modal approach that included 6 specific upper-body postures. Those implementations use a small set of gestures, whereas others have made use of a reusable database with about 500 annotated gestures (Vilhjálmsson et al., 2007; Pelachaud et al., 2010). Those authors also describe a Behavioral Markup Language that allows virtual and physical presenters to use and combine these gestures. Despite sharing the goal of non-verbal communication with robots, most studies define gesture sets which are either specific to the task or to the robot. While the Behavioral Markup Language aims to overcome the latter, the iconicity and cultural-dependence of most gestures makes it is difficult to see how the implementation of task-specific gestures can be easily generalized.

2.1.2 Schematic Movement

We can also explore commonalities among gestures with regard to their embodied semantics. Cienki (2013a) argues that gestures ground the cognitive model of situated speakers in their physical environment. The schematic nature of certain movements across different gesture types has been related to image-schematic structures. These are recurring cognitive structures shaped by physical interaction with the environment (Lakoff, 2008; Johnson, 2013), and can be observed not just in verbal but in non-verbal communication (Cienki, 2013b; Mittelberg, 2018). Johnson provides the example "*Let out your anger*" (Johnson, 2013). Here, anger, a metaphorical "fluid" housed in the body, is said to be released from its container. As shown in (Wicke and Veale, 2018c), such schemas can be used to depict causal relations in embodied storytelling with robots. Not only does the theory of image schemas provide a *Conceptual Scaffolding* (Veale and Keane, 1992) for narrative processes, it also provides an algebraic basis for modeling complex processes and situations (Hedblom, 2020). In this way, simple schemas can be used as primitive building blocks of larger, more complex structures (Veale and Keane, 1992; Besold et al., 2017). For example, Singh et al. (2016) describe a playful co-creative agent that interacts with users by classifying and responding to actions in a 2D virtual environment. These authors train a Convolutional Neural Network on schematic movements so that it can classify inputs as, for instance, *Turn*, *Accelerate* or *Spin*. Moving from a two-dimensional plane to three-dimensional embodied space allows us to combine gestures with other body movements that extend beyond gesture space. Those extended movements can also tap into our stock of embodied schemas to support a metaphorical understanding of physical actions.

2.1.3 Body Movement

An advantage of the schematic approach is that a small set of robotic movements can produce a large number of useful combinations. For our current purposes we define just two types of bodily movements:

- **Spatial:** Movement along one axis
- **Rotational:** Rotation around one axis

Each type of movement fulfills a physical function: Spatial movement changes the position of the robot in space, while rotational movement changes the direction the robot is facing. It is known from the early studies of Heider and Simmel (1944) that even simple movements can lead an audience to project intentional behavior onto inanimate objects, to perceive emotion where there is only motion. A study by Nakanishi et al. (2008) shows that even minimal movement on one axis of a robot-mounted camera increases one's sense of social telepresence. Implementing rotation and directional movement in a museum robot, Kuzuoka et al. (2010) show that a robot's rotation can influence the position of a visitor, and that full body rotation is more effective than partial, upper-body rotation. Nakauchi and Simmons (2002) have investigated literal spatial movement in the context of queuing in line, and consider relative positioning in line as a parameter for achieving optimal, socially-accepted movement. Our focus here is on the metaphorical potential of bodily movement in a robotic context that must speak to human emotions. **Table 1** provides examples of how schematic constructions, implemented simply with robots, can convey intention and emotion. Of course, even the *metaphors we live by* (Lakoff and Johnson, 2008) can brook exceptions. For example, UP may generally signify good, and DOWN bad, but we want a fever to go down, and do not want costs to go up. This observation also applies to the schemas presented in **Table 1**. There are some situations where moving away increases emotional closeness, and moving closer decreases it, as when e.g., the former signifies awe and great respect, and the latter signifies contemptuous familiarity. As with all powerful schemas, we believe the benefits of generalization outweigh the occasional exceptions.

Following Falomir and Plaza (2020), who argue that primitive schemas like these can be a source of creative understanding in computational systems, we believe that simple schemas can be reused across creative applications of robotic movement, to connect the semantics of the task with the movement of the robots. Each movement may carry a unique semantics for different tasks, yet build on the same schematic basis. For example, the choreography of dancing robots can be synchronized using the same basic motions (back, forth, left, right). The dance can reflect abstract concerns through metaphorical motions, as when robots dance in a circle to reflect a repeating cycle of events. Likewise, in a storytelling context, actors can strengthen a perceived bond by moving closer together over the course of a story, or weaken and break that bond by gradually moving apart.

2.1.4 Limitations by Context

"Space" is a very general notion that can be understood in different ways in different performative contexts. For instance, our understanding of the movements of fellow pedestrians on the street is subtly different from our understanding of actors pacing about a designated stage. Even stages differ, and a proscenium arch can frame the action in a way that encourages a different kind of dramatic interpretation than a stage that is not so clearly divided from the viewing gallery. So our perception of how space is framed can influence our construal of meaning within that space (Fischer-Lichte and Riley, 1997). Just as the physical stage frames our conception of space, physical actors frame our notions of gesture and locomotion. We make a simplifying assumption in this work that our robots exhibit comparable degrees of freedom to an able-bodied human actor, but this need not be the case, Different robot platforms presuppose different kinds of movements, and support different degrees of physical verisimilitude (see **Section 6.3**). While we accept the limitations of our current platform, the anthropomorphic Nao, and choose our actions and schemas to suit these limitations, other robot platforms may afford fewer or greater opportunities for embodied meaning-making.

2.2 Exploring Meaning in Movement

Robots do more than stand in for the characters in a story. Their performances should convey meaning that augments that of the spoken dialogue and narration. When we speak of the semantic interpretation of movements and gestures, it is tempting to ground this interpretation in a componential analysis, and ask:

TABLE 1 | A listing of image schemas with their robotic realizations, with additional potential for metaphorical meaning. Each row contains a schema and its inverse.

Schema	Movement	Metaphorical Meaning	Schema (Reverse)	Definition	Metaphorical Meaning
NEAR	The robot is moving near another robot or object	There is an interest or sympathy towards the robot/object	FAR	The robot is moving further away from the robot/object	There is a growing dis-interest or disliking to-wards the robot/object
FRONT	The robot is moving or turning in front of itself or another ro-bot/object	The robot is actively en-gaging with the other ro-bot/object	BACK	The robot is moving or turning to the back of itself or another robot/object	The robot is actively dis-engaging with the other robot/object
UP	The robot is moving upwards	The robot is displaying some super-iority over the other robot/object	DOWN	The robot is moving down or downwards	The robot is displaying some infer-iority over the other robot or object

what are the components of gestures and other movements that convey specific aspects of meaning? In sign language, for instance, signs have a morphemic structure that can be dissected and analyzed (Padden, 2016). But gestures are not signs in any sign-language sense, and cannot usually be dissected into smaller meaningful parts. Indeed, signers can use gestures with sign language, just as speakers use them with spoken language (Goldin-Meadow and Brentari, 2017). Moreover, there is some neuro-psychological evidence that speech-accompanying gestures are not processed by language-processing mechanisms (Jouravlev et al., 2019). Our gestures give additional context to speech, while speech gives a larger context to our gestures. They add meaning to language (Kelly et al., 1999; Cocks et al., 2011) while not strictly constituting a language themselves. Some gestures indicate that a speaker is looking for a certain word [these are called *Butterworth* gestures by McNeill (1992)], and so serve a meta-communicative function. Likewise, a speaker can produce many kinds of unplanned movement while communicating with language, such as tugging the ear, scratching the head or waving the hands, and although an entirely natural part of embodied communication, we do not seek to replicate these meta-communicative gestures here. Rather, our focus is on gestures that communicate specific meanings, or that can be used to construct specific metaphors.

2.3 Creative Robots in Other Performative Contexts

Story-telling is just one performative context in which robotic actors use space and movement to convey meaning. For instance, robots have been used for improvisational comedy. *ImprovBot* (Rond et al., 2019) collaborates with human actors in ways that require it to spin around, move in circles, or move forward, backward, and sideways. Similarly, the robotic marimba player *Shimon* (Hoffman and Weinberg, 2011) recognizes the gestures of a human collaborator, and uses a schematic understanding of those gestures to make corresponding music-making decisions. A robot artist that creates photo montages and digital collages by interacting with a human user is discussed in Augello et al. (2016a). The robot takes its cue from a variety of information sources, one of which is the posture of its user. A design for a robot artist that interacts with a human user in a therapeutic setting is sketched in Cooney and Menezes (2018). Again, the aim is integrate a range of cues, both verbal and physical, from the human into the robot's physical actions. Just how well robots like these mesh with their collaborators, whether human or artificial, is the basis of the interactional "fluency" explored by Hoffman (2007).

When creative robots use motion to convey meaning, we expect them to aim for more than the "mere execution" of a literal script. Augello et al. (2016b) explicitly make the latter their goal, in the context of a robot that learns to dance in time to music. Another dancing robot system, that of Fabiano et al. (2017), chooses its actions to match the schematic drawings of dance movements shown to it by a collaborator.

Robot actors on a stage can be likened to human actors in a stage play, or to animated cartoon characters. In each case, however, the artifice succeeds to the extent that movements are considered natural. Laban movement analysis (LMA), which allows one to characterize the effort required for different bodily movements, in addition to modeling the body's shape and use of space, has been used by Bravo Sánchez et al. (2017) to support natural robot movements in short plays. Robots can enact artificial stories generated by an AI system, or they can interactively enact a human-crafted story. The *GENTORO* system of Sugimoto et al. (2009) does the latter, to encourage story-telling in children by combining robots and handheld projectors. A story-telling (or story-enacting) robot can be a physical presence, or a wholly virtual one, as in Catala et al. (2017). Nonetheless, results reported in Costa et al. (2018) show that embodied robots garner more attention and engagement that virtual ones.

As noted in (Augello et al., 2016b), a performer should do more than merely execute a script. Rather, it should interpret that which it sets out to perform, in whatever context – conversation, theatre, dance – it is designed to do so.

One of the first computational storytelling systems to consider context was *Novel Writer* Klein et al. (1973), a system for generating short tales of murder in a specific context (a weekend party). Simulation is used to determine the consequences of events as shaped by the chosen traits of the killer and his victim. Changing these traits can alter the simulation and produce different plot outcomes. Story-telling systems can obtain and set these plot-shaping traits in a variety of

TABLE 2 | Examples of gestures: *Iconic*, *Deictic*, *Metaphoric*, *Cohesive* and *Beats* by a robot. Each gesture is ascribed a general property, along with its hardware requirements.

Gesture Type (*Example*)	Depiction (*Example*)	Req. Hardware	Properties
Iconic (*Drive*)		Arms, Hands	Obvious
Deictic (*Point*)		Pointing limb	Relational
Metaphoric (*PUOH*)		Arms, Hands	Metaphorical
Cohesive		Limb	Global
Beats		Limb	Local

ways, both direct and indirect. A robot storyteller can, for instance, obtain personal traits from its users, by asking a series of personal questions that are shaped by its own notions of narrative (Wicke and Veale, 2018a). Those questions, and the answers that are provided, then shape the generated story, and provide a context for the audience to understand the actions of the performer. Basing a story on a user's own experiences is just one way of providing a clear interpretative basis for the performer's actions on stage. The more general approach provided here seeks to instead ground the performer's choices in a user-independent model of how characters are affected – and are seen to be affected – by the cumulative actions of the plot.

3 PERFORMANCE FRAMEWORK FOR ROBOTIC ACTORS

3.1 Technical Description of Movements

The *Performance Framework* is applicable to a variety of embodied performance types that include robots, such as dancing, storytelling, joke telling and conversation. While those tasks impose unique requirements for hardware and software, the framework provides a unified conceptual perspective. The next sections present the framework, and explain its terminology and its syntax for describing movement. We start with a technical description of the specific movements that can be derived from the conceptual organization of movement types. As in **Section 2.1**, we address gestures first, followed by body movements.

3.1.1 Gestures

The properties of the five types of gestures described in (**Section 2.1.1**) are listed in **Table 2**, along with illustrative examples of each gesture. We illustrate each gesture in its most iconic form, with the exception of the *Beats* and *Cohesive* gestures, since these are always specific to the temporal context in which they appear. As noted earlier, the *Iconic* and *Metaphoric* gestures can use the same movements to convey different meanings in different contexts.

Iconic: An iconic gesture has an obvious meaning, since an icon can clearly substitute for what it is supposed to represent [see Peirce (1902) and Mittelberg (2019); the latter provides a thorough linguistic discussion of signs, icons and gestures]. These icons of physical actions are schematic by nature, insofar as they enact patterns of embodied experience (Mittelberg, 2019). We therefore attribute the property of *Obviousness* to the *Iconic* gestures, since they make meanings more explicit, and leave little room for alternate interpretations. **Table 2** presents an example of a robot steering an invisible vehicle to iconically depict the act of *driving*. Despite their obvious iconicity, many iconic gestures can be culturally-specific. As discussed in **Section 2.1.1**, gestures that are obvious to the speakers of one language may be confusing, misleading and far from obvious to members of a different cultural or linguistic grouping.

Deictic: Since pointing gestures refer to spatial/physical landmarks, we ascribe to *Deictic* gestures the property *Referential*. The technical implementation of such a gesture requires a limb, ideally an arm, that can point at the target reference. It is also beneficial if the pointing is further supported

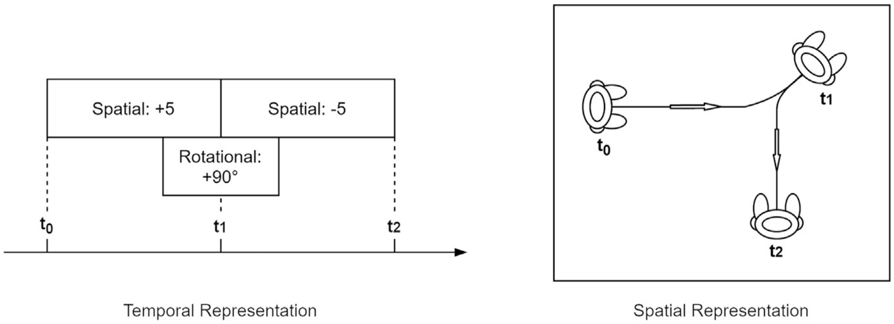
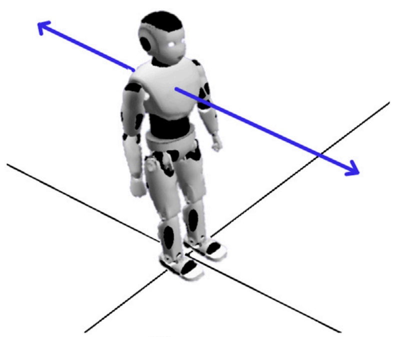
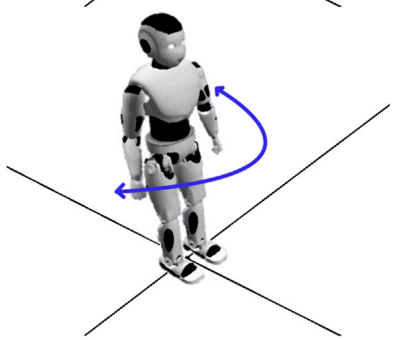


FIGURE 2 | Combination of Spatial and Rotational Movement in two representations. Left: Spatial movement can co-occur with rotational movement. Different directions can be achieved by combining rotations with positive or negative spatial moves. Right: The result of parallel spatial and rotational movement is a curved walk. Here t_0 is the robot's position prior to movement. t_1 is the point when the robot shifts from forward movement to backward movement, whilst completing half of the 90 degree rotation.

TABLE 3 | Depiction of the body movements *Spatial* and *Rotational* with their corresponding name, depiction with a physical robot, movement vector and properties in the respective columns.

Movement	Depiction	Transformation Matrix	Properties
Spatial		$S_x(\omega) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \omega & 0 & 0 & 1 \end{bmatrix}$	• Global • Relational • Summative • Additive • Persistent
Rotational		$R_z(\alpha) = \begin{bmatrix} cos(\alpha) & -sin(\alpha) & 0 \\ sin(\alpha) & cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix}$	• Relational • Obvious • Summative • Additive • Persistent

by the head or gaze direction of the robot (Clair et al., 2011). As with *Iconic* gestures, *Deictic* gestures also overlap with metaphorical gestures in different contexts (e.g., pointing ahead of oneself to signal a future event). **Table 2** shows the example of a robot pointing ahead with its arm.

Metaphoric: *Metaphoric* gestures, labelled *Metaphorical* in **Table 2**, are the most challenging to implement since their intent must be discerned *via* a mapping from literal to non-literal meanings. Yet, as a consequence of this mapping, metaphorical gestures also open new possibilities for creativity within the system. An example of the creative use of metaphorical gestures is provided in **Section 4**.

Cohesives: Cohesive gestures are dependent on their context of use, and require careful timing. Whether a shaking of the fist, a circling of the finger or a turning of the wrist, such movements only make sense in a narrative if they are used coherently. Coherent usage aids discourse comprehension and allows audiences to construct a spatial story representation (Sekine and Kita, 2017). Moreover, *Cohesives* can strengthen our grasp of the whole narrative if they are used recurrently to reinforce persistent or overarching aspects of the plot. We therefore ascribe the attribute *Global* to these gestures.

Beats: *Beats* are just as context-dependent as *Cohesives*, but lack the latter's global influence, as they are relevant to one-off events

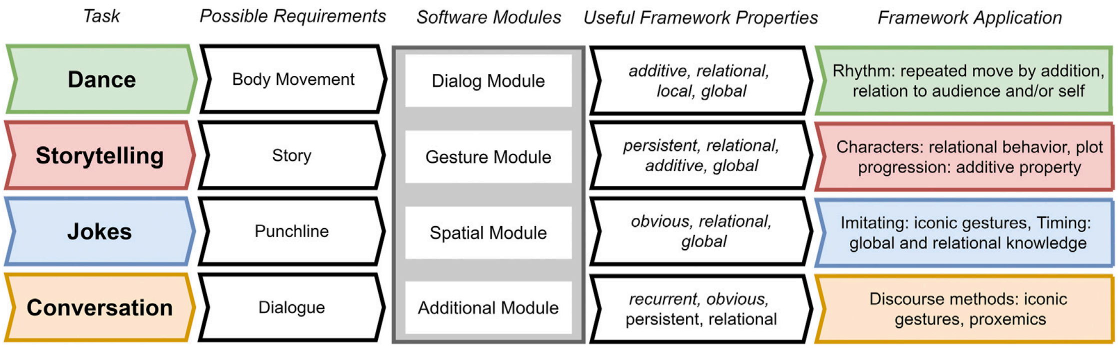


FIGURE 3 | Diagram of four possible tasks that can be accounted for by the framework. From left to right, each task poses some prototypical requirements for the corresponding software module, which can be addressed by the performance framework properties.

only. Since the movement itself is less relevant than its timing and its context of use, no concrete example is offered in **Table 2**. In opposition to the *Cohesives*, we label these gestures as *Local*. Note that a gesture that is considered a *Beat* in one task domain, such as story-telling, might serve a global role in the synchronization of movement in another, such as dance. In that case, the gesture would be labelled *Global* in the latter context.

Since the gesture types illustrated in **Table 2** do not constitute an exhaustive list, some additional properties may need to be included in the future. For example, some gestures are performed with two hands and so can exhibit relational properties in the way that each hand, representing a character, relates to the other (Sowa and Wachsmuth, 2009). We can also consider the naturalness of the gesture, as this is an important property for HRI and a common basis for assessing any computational model that uses gestures (Salem et al., 2012; Huang and Mutlu, 2013). However, we might also view naturalness – as we do here – as an emergent property of the implementation, rather than as a constitutive property of the performance framework itself.

3.1.2 Body Movement

Spatial Movement describes a simple trajectory of an agent along one axis. This core movement requires the agent to possess a means of locomotion, such as wheels or legs. In its basic form, spatial movement in one direction is a transformation of the positional coordinates in one variable:

$$\vec{x} = (\omega, 0, 0) \qquad with\ \omega \in \mathbb{R} \tag{1}$$

The corresponding translation matrix is given in **Table 3**. This movement is compatible with all other movement types. Combined with rotational movement, it covers all directions on the 2D plane. When the mode of locomotion allows for it, the vector can be positive or negative. This kind of movement can exhibit the following properties:

- **Global:** The moving body affects the relative proximity, shared references and spatial configuration of all agents in a performance, and so has implications for the performance of the narrative as a whole.
- **Relational:** The movement has implications for other agents on the stage since an absolute change in position for one actor also changes its position relative to others.
- **Summative:** The movement of an actor into its resulting position summarizes, in some general sense, the history of past actions up to this point.
- **Additive:** A movement compounds a previous action to achieve a perceptible cumulative effect.

TABLE 4 | Possible combinations of movement types. *comb.* (green) are **combinable** movements, *restr.* (yellow) are combinations that are only possible to some **restricted** extent and *excl.* (red) are movements that are mutually **exclusive**.

Movement Type	Spatial	Rotational	Iconic	Deictic	Metaphoric	Cohesive	Beats
Spatial	comb.	comb.	restr.	restr.	restr.	comb.	comb.
Rotational	comb.	comb.	restr.	restr.	restr.	comb.	comb.
Iconic	restr.	restr.	comb.	comb.	excl.	comb.	excl.
Deictic	restr.	restr.	comb.	comb.	comb.	comb.	excl.
Metaphoric	restr.	restr.	excl.	comb.	comb.	comb.	excl.
Cohesive	comb.	comb.	comb.	comb.	comb.	comb.	excl.
Beats	comb.	comb.	excl.	excl.	excl.	excl.	comb.

- **Persistent:** A movement has a lasting effect on an actor or its the physical relationship to others.

The property *Obviousness* is not attributed here, since actors (robotic or otherwise) can move in space without necessarily conveying meaning. Some movements help a speaker to communicate while being uncommunicative in themselves, as when an actor steps back to maintain balance, or moves their hands in time to their words as they speak. In contrast, it is hard to perceive a rotational movement as unintentional, since rotation carries such an obvious, iconic meaning. Thus, while the property of *Obviousness* is not wholly context-free, it is sufficiently robust across contexts to earn its keep in a performative HRI system.

Rotational Movement This movement requires an actor to possess some form of rotational joint, so that it can rotate around one axis. While a humanoid robot can simply turn its head or torso, this kind of rotational movement requires full body rotation. In some cases, rotation is only possible in combination with spatial movement. For example, some bipedal robots cannot rotate on the spot, and need to walk in a curve to achieve full rotation. The rotation around one axis is given by the transformation matrix in **Table 3** as $R_z(\alpha)$ (with α as degree of rotation). This kind of movement can exhibit the following properties:

- **Obviousness:** When the movement achieves the iconic action of turning away from, or turning toward someone else, this iconicity deserves the label *Obvious*.
- **Relational:** The rotation has implications for other agents on the stage since an absolute change in orientation for one actor changes its relative orientation to other agents.
- **Summative, Additive and Persistent:** These properties hold the same meanings for rotational movements as they do for spatial movements.

3.2 Combinations of Robotic Movement

Defining the basic types of movement and their properties provides a foundational set of movements that can be implemented for different kinds of robots. Basic movements can be considered primitive actions in a performance system, whose possibility space is the space of their possible combinations. Gestures can be combined with whole body movements (spatial and rotational) to produce complex behaviours. The individual movements themselves are not creative – many are iconic, and highly familiar – but the mapping from narrative to physical action does allow for metaphor and for other creative choices (Boden, 2004). The example combination provided in **Figure 2** shows a forward movement followed by a backward movement, paralleled by a rotational movement during the transition. The resulting performance (see **Figure 2** *Spatial Representation*) is the sum of its parts, and fosters audience interpretation of the performer's behaviour. This is where the properties *Summative*, *Additive* and *Persistant* come into play.

An embodied performance can draw on all available movements and all possible combinations of such. **Table 4** presents a combination matrix showing possible combinations, mutually exclusive movement types, and restricted combinations. The group that is least conducive to interaction with others is the *Beats*. Due to their *local* property, these are grounded in a specific narrative moment, which does not permit metaphorical, iconic or deictic displays. This momentary status also strongly prohibits combinations with *Cohesives*. In short, only *Beats* can combine with *Beats*. Nonetheless, *Beats* can be performed during spatial or rotational movements, as this does not change their function. In fact, spatial and rotational movements can be combined with all other movement types, as well as with themselves. However, a spatial or rotational movement during an iconic or metaphoric gesture can cause positional changes that affect the gesture, while deictic gestures are also sensitive to any referential changes of position. For example, pointing while walking is a much more restrictive task than either alone, since the target of the reference might move behind the performer.

By definition, iconic and metaphoric gestures exclude each other. As with a change of context, a gesture's obviousness can be exchanged for a metaphorical interpretation, but a combination of iconic and metaphoric gestures must be sequential, not parallel. Likewise, the *Cohesives* can be combined with any other movement type except for the *Beats*, since these groups have opposing *global* and *local* properties.

3.3 Performance Framework at Work

Figure 3 depicts four example tasks, the requirements of each, and the applicability of the framework to each instance of the task. The framework is designed to meet the demands of these different tasks. When the *Software Modules* for a task depend on the choice of performing agents (e.g., embodied/non-embodied, single/multiple), the properties needed to support an appropriate conceptual response are given.

Different performances types can place varying emphases on the meanings of any given movement. Dancing is an expressive act which aims to convey themes and emotions through the use of the entire body. While **Figure 3** lists only *Body Movement* as a necessary requirement for dance, dancing can have other requirements in context, e.g. single or multiple bodies which can - but do not need to - move synchronously. Dance types can range from the highly coordinated to the highly improvised and relatively uncoordinated. As shown in **Figure 3**, rhythm can, for example, be achieved with a repetition of movements. However, while rhythm and synchrony are listed as prototypical requirements of a dance task, these are neither necessary nor sufficient for dance, and this point applies more generally to all performance types, from dance to storytelling to joke-telling and casual conversations. In robotic dance, complex relational movements and motion dynamics are at play, which may or may not exhibit synchrony and rhythm (LaViers and Egerstedt, 2012; LaViers et al., 2014; Thörn et al., 2020). Within the *Performance Framework*, additive, relational, local and global properties can be identified, and, in cases where it is required, rhythm can be achieved by adding repeated movements, just as synchronized movement can be realized in terms of global and relational additions. Ultimately, *Body Movements* is a flexible requirement which should always acknowledge the diversity of bodily capabilities across humans and across robots, making it all the more important that each possible requirement is appropriately integrated on the software level.

For storytelling, Wicke and Veale (2020b) have shown that movement, gesture and relative positioning play an important role in enacting a story well. These requirements can each be met using movements with persistent, relational, global and additive/summative properties. When actors undergo changes in their physical spaces that mirror the changes undergone by characters in a narrative space, metaphorical schematic movements and gestures can depict plot progression and character interrelationships.

Certain performance types, such as stand-up comedy, place a greater emphasis on timing than others (Vilk and Fitter, 2020). When timing is key, spatial movements may be subtle and minimal (Weber et al., 2018), making the properties *obvious*, *relational* and *global* all the more important. For example, the timing needed to land a punch line requires global and relational knowledge of the performance as a whole, while the use of iconic gestures throughout can increase the effectiveness of the performance.

Lastly, conversational agents make use of various discourse strategies that can be enhanced by the use of iconic and deictic gestures. The latter are especially useful in maintaining shared attention and awareness, by mirroring movement in a topic space with movement in physical space (Jokinen and Wilcock, 2014).

4 AN INTERPRETATION FRAMEWORK FOR STORYTELLING ROBOTS

By meaningfully connecting plot actions to movements, the *Performance Framework* allows a performer to pick its movements to suit the action (x) at hand. More formally, $C(x) \mapsto E(x)$ denotes the mapping of an action x as it is represented in the conceptual domain C of stories to its expressive realization in the embodied, physical world E. For example, the *insult* action can be expressed with an iconic gesture in which an actor "flips the finger" to another actor. That other actor may show that they feel disrespected by moving their head slightly backwards. In this case, $C(insult) \mapsto E(insult)$ because the actors physically express the *insult* action that the plot calls for. However, each actor should take into account the current state of the story, and their residual feelings that carry over from earlier actions. If we denote this state of the story as S, then the performers consider the mapping $S(C(insult)) \mapsto S(E(insult))$. In a story space with x_N possible actions, the general form of this mapping is $S(C(x_N)) \mapsto S(E(x_N))$.

Skilled actors are able to interpret an action within the context of the unfolding story, so we also need a complementary *Interpretation Framework* to allow performers to interpret each action in context. Suppose character A has supported B in some way, or confided in B, or defended B, and B then responds by insulting A. Viewed in isolation, the insult should make A feel disrespected, and even a little attacked, so it would be appropriate to embody this event as $C(insult) \mapsto E(insult)$. However, given the earlier events which make this insult all the more shocking, it would be even more appropriate, from A's perspective, to enact $C(insult) \mapsto E(attack)$, since *attack* carries more shock value than *insult*. Each performer brings a different interpretation to bear on the same plot action. So while B interprets the *insult* action directly, A interprets it as *attack* action. The result is a performative blend of the two enactments. B enacts its agent role in the *insult* while A enacts its patient role in the *attack*. That is, while B enacts the event *via* the mapping $S(C(insult)) \mapsto S(E(insult))$, A uses the mapping $S(C(insult)) \mapsto S(E(attack))$. The more general form of A's interpretation is $C(x) \mapsto E(\overline{x})$. It is the task of the *Interpretation Framework* to provide the mapping mechanisms for interpretations such as these.

4.1 The Representation of Gestures Within the Framework

Wicke and Veale (2018b) define one-to-many mappings from plot actions to gestures and movements, from which performers can choose an appropriate but context-free enactment at random. The purpose of the performance framework is to transform this choice from a purely disjunctive one to a choice based on interpretation in context. To this end, each gesture must be understood by the system as more than a black box motor script. So in addition to duration information, a schematic classification as given in **Table 1**, and the properties given in **Table 2**, we must give the framework an emotional basis for making real choices.

Our database of gesture representations is available from a public repository[1]. More than 100 movements are currently stored and labelled for use in embodied storytelling. Each is assigned a unique name that describes the movement briefly. This name aims to be as telling as possible in just a few words, while a longer free-form description is as explicit and detailed as possible. For example, the movement named "strike down" has the description "right arm squared angle lifted above shoulder, quickly striking down with hand open." This movement, which takes approximately $2\frac{1}{2}$ seconds to execute, is labelled as a schematic *down* movement. This motion is not associated with rotational or spatial movement, and its possible uses as an iconic or metaphorical gesture depend on the narrative context in which it is performed.

As noted in (Wicke and Veale, 2018b), the existing disjunctive mapping from plot actions to physical actions is further labelled with an appropriateness label, since some gestures are more obviously suited to their associated plot actions than others. For example, the action *disagree_with* can map to either of the gestures "shaking the head" or "shaking the head, raising both arms and turning away." In this case, both gestures are equally appropriate for the action. For another action, however, such as *contradict* or *break_with*, the latter is more appropriate than the former. Three distinct appropriateness levels – *high*, *medium* and *low* – are used to qualify the mappings of actions to gestures. This suggests a very practical motivation for metaphor within the system: the mapping $C(x) \mapsto E(\overline{x})$ is preferable to $C(x) \mapsto E(x)$ when $E(\overline{x})$ offers a more

[1]https://osf.io/e5bn2/?view_only=2e30ee7e715342d59c371b5d30c014e0

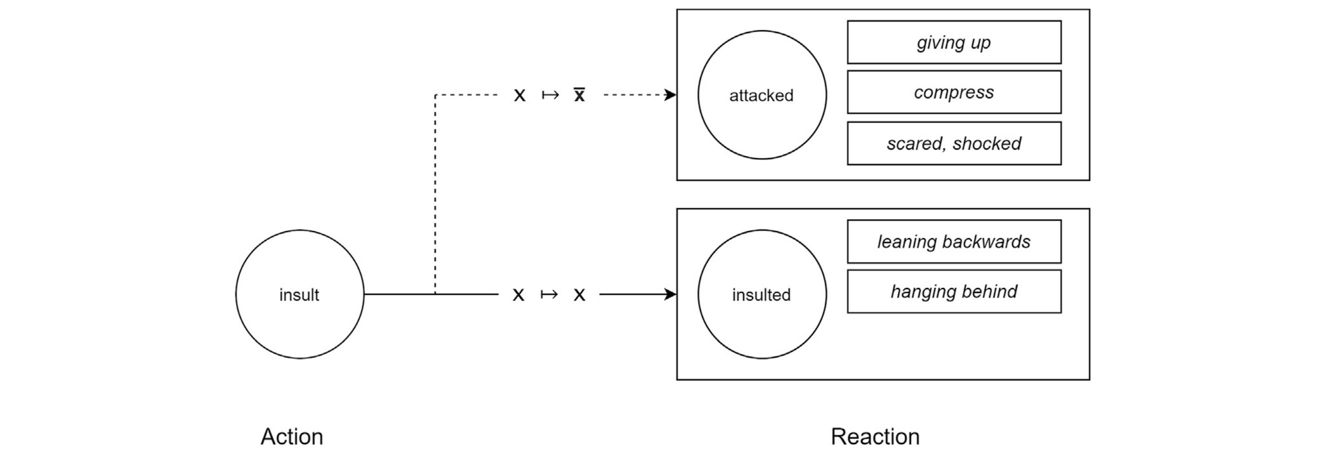


FIGURE 4 | Depending on the narrative context, the action A *insult* B can cause B to feel "insulted" ($C(x) \mapsto E(x)$) or to feel "attacked" ($C(x) \mapsto E(\bar{x})$). B's emotional response will then dictate the actor's physical reaction on stage.

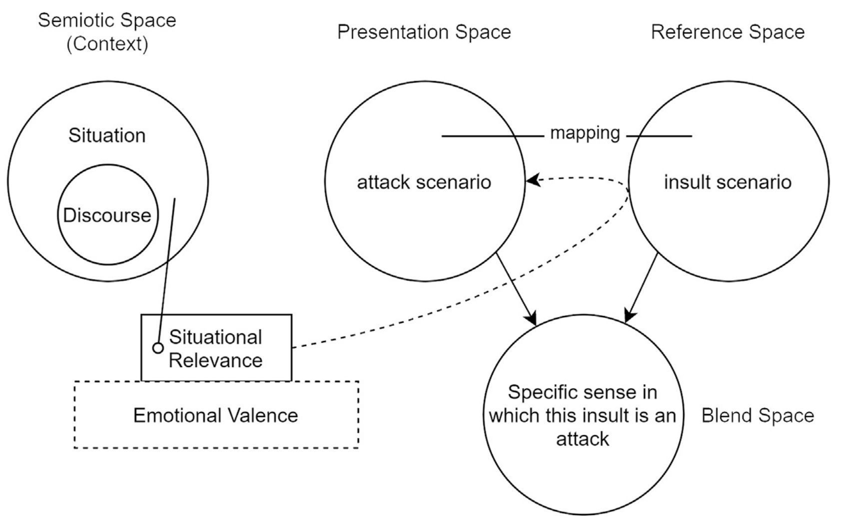


FIGURE 5 | Selective projection in a situation in which an insult is interpreted as an attack. Situational relevance is determined by the emotional valence of the current action. Figure based on context-dependent blending theory by Brandt and Brandt (2005), and adapted from Li et al. (2012).

appropriate enactment for x than $E(x)$. In general, metaphor will be motivated by a mix of concerns, from the practical (does this action have a vivid enactment that really suits it?) to the expressive (does this action adequately capture the feelings of the moment?). Notice that in each case, however, metaphor hinges on questions of expressive adequacy, and the question of whether the systems knows of a better way to communicate what it wants to say.

4.2 Selective Projection for Creative Interpretation

An embodied performance of a story is a careful presentation of story elements – plot, character, emotion – in a physical space. As such, performers project elements from the story space, a space of words and concepts, into the presentation space, a space of gestures and movements and spoken dialogue. The performers are themselves, with their own physical affordances and limitations, *and* the characters they play. In the terminology of Fauconnier and Turner (1998), Fauconnier and Turner (2008), the performance is a *conceptual blend*.

Turner and Fauconnier's *Conceptual Blending Theory* has previously been used to model stories in a computational setting (Li et al., 2012). The basic theory has been extended by (Brandt and Brandt, 2005) to incorporate additional spaces that are especially relevant here, such as a reference space (for the underlying story), a context space (specifying situations within the story, and discourse elements relating to those situations), and a presentation space in which story elements are packaged and prepared for a performance.

Consider again the example story in which character B insults A after A has shown favor to B, perhaps by praising, aiding or defending B. In the reference space this plot action is literally captured as *B insult A*. As mediated by the context space, however, which brings both situational relevance and a discourse history to the interpretation of events, B views this insult as an attack, and so the action is instead represented in the presentation space as *B attack A*. Since the performers take their stage directions from the presentation space, A will move, gesture and speak as though the victim of an actual attack. So, when B performs a "giving the finger" gesture to A, A will do more than lean back in disappointment – the standard response to an insult – it will step away with its arms extended in a defensive posture. This construal of events by A and B, the first of three scenarios unpacked below, is illustrated in **Figure 4**.

Scenario 1: An insult delivered in some contexts can surprise more, and wound more, than in others. The standard response, which entails a literal mapping from the reference to the presentation space, is $S(C(insult)) \mapsto S(E(insult))$. However, in a story state S that makes the insult seem all the more severe, $S(E(insult))$ may equate to $E(attack)$, to produce the non-standard mapping $S(C(insult)) \mapsto E(attack)$. In that case, it is not the embodied response to an insult, $E(insult)$ that is enacted by the insult's target (leaning back, with head down) but $E(attack)$ (stepping back, arms outstretched defensively).

Scenario 2: A performer whose character, A, praises the work of another, B, might enact a show of "praise" with a clap of the hands or a nod of the head. This is the standard response in a story context where praise is literally interpreted as praise, that is, $S(C(praise)) \mapsto S(E(praise))$. However, if the context indicates that A has strong grounds to respect and feel inspired by B – perhaps B rescued A in the previous action – then $S(E(praise))$ may be interpreted in this light to produce a stronger reaction than praise. As such, $S(E(praise))$ might equate to $E(worship)$ and the performer playing A will bow accordingly.

Scenario 3: A succession of actions that reinforce the same emotional response in a character can lead to a character feeling and expressing that emotion to a higher degree, shifting its embodied response from the standard interpretation to a heightened, metaphorical level. Suppose the story concerns character A treating character B as a lowly minion. A overworks and underpays B, taking advantage of B at every turn. If A should now scold B, B may interpret $S(E(scold))$ as $E(whip)$, or interpret $S(E(command))$ as $E(enslave)$, and finally interpret $S(E(fire))$ as $E(release)$. Interpretative performance allows for a shadow narrative to play out in physical actions as the literal narrative is rendered in speech.

In each scenario, the situated actor uses contextual information to interpret the current plot action, in light of previous actions, and chooses to accept the scripted action ($x \mapsto x$) or to take a metaphorical perspective ($x \mapsto \overline{x}$) instead. An alternate construal, such as construing an insult as an attack, or an act of praise as an act of worship, changes the physical enactment of the action in the script. Notice that when an alternate enactment is chosen, the dialogue associated with the scripted action is still used. The combination of one action's gestures with another action's dialogue adds further variety to the blend, while also helping to foster understanding by the audience. Gestures are dramatic on a physical level, but dialogue carries a more explicit semiotic content. Even when the performers choose to be metaphorical, the performance remains grounded in some literal aspects of the script. This grounding is rooted in the assumption that audiences are capable of fully comprehending the narration and dialogue of the script. When this is not the case, gestures and other non-verbal cues become an even more important channel of communication.

A blending interpretation of *Scenario 1* is illustrated in **Figure 5**, further adapting the treatment of Brandt and Brandt (2005) that is offered in Li et al. (2012). Notably, situational relevance is informed by the *Emotional Valence* of the situation, the calculation of which we consider next.

4.3 Emotional Valence in Story Progression

Veale et al. (2019) have shown how the actions, characters and structural dynamics of a story can influence the performers' reactions so as to elicit a comedic effect in a performance by robots. In that approach, the logical structure of the narrative – in particular, whether successive actions are linked by "but" or "then" or "so" – provides a reasonable substitute for an emotional interpretation of the action, so that performers know when to act surprised, or can infer when an audience might be getting bored (e.g., because the plot lacks "but" twists) or confused (e.g., because it has too many "but" twists). To go deeper, we must augment this structural perspective with an emotional perspective, so that performers can grasp why certain actions are linked by a "but" and not a "so". We begin by situating each role (A and B) of every possible action in a plot (*Scéalextric* defines more than 800 different actions) on the following four scales:

disappointed ← A → inspired	disappointed ← B → inspired
repelled ← A → attracted	repelled ← B → attracted
attacked ← A → supported	attacked ← B → supported
disrespected ← A → respected	disrespected ← B → respected

These emotions are chosen to suit the action inventory of a story-telling system like *Scéalextric*. Other emotional scales may be added as needed to suit other tasks, such as dance (Camurri et al., 2003). The story-telling system draws from a knowledge base of over 800 actions, which can be causally connected to create stories that exploit tropes and other common narrative structures. Each story revolves around two central characters and a retinue of secondary figures (partners, spouses, friends, etc.). The most common themes elicit feelings of trust, respect, admiration and cooperation about and between those characters.

For example, the *insult* action associates a strong sense of being disrespected with the B role. When A insults B, we expect B to feel very disrespected (or negatively respected). Similarly, the *worship* action associates a strong sense of being inspired with the A role, and a strong sense of being respected with the B role. Conversely, the *surrender to* action associates a negative sense of being attacked (and so a positive sense of being supported) with the B role, because A is no longer an active threat to B. Viewed individually, each emotional setting can be compared to that of the previous action, to determine how much change has been wrought by the current plot turn. It is this change that explains why certain transitions warrant a "but" and others warrant a "so" or a "then". It can also motivate why an insult can come as a surprise to a character, and feel more like an attack.

The four emotional scales can also be viewed in the aggregate, to determine an overall valence for the current action from a given role's perspective, or to determine an overall shift in valence from one action to the next. We calculate the valence of a role in an action α_i as the total valence across all emotional scales for that role in that action. See **Eqs 2**, **3** for the valence of the A and B roles in α_i. A positive valence for a role indicates that a character in that role should experience a positive feeling when the action is performed; conversely, a negative valence suggests a negative feeling for the action.

$$valence_A(\alpha_i) \leftarrow inspiration_A(\alpha_i) + attraction_A(\alpha_i) + support_A(\alpha_i) + respect_A(\alpha_i) \tag{2}$$

$$valence_B(\alpha_i) \leftarrow inspiration_B(\alpha_i) + attraction_B(\alpha_i) + support_B(\alpha_i) + respect_B(\alpha_i) \tag{3}$$

A character is a persistent entity in a narrative, one that moves through the plot from one action to the next. The current valence of a character is a function of the valence of the role it plays in the current action, and of the valence of its roles in previous actions, with the current action making the greatest contribution. Previous actions have an exponentially decaying effect based on their recency. If $0 < \beta < 1$ specifies the weight given to the current action, the contextual valence of the characters filling the A and B roles is given by **Eqs 4**, **5** respectively. We assume a fixed decay rate, while acknowledging that certain events might have a stronger and more lasting impact on perceived valence than others. It remains to be seen in future work whether this simple one-size-fits-all approach needs to be replaced with a more variable, local solution. For now, we continue to view this simplicity as a virtue.

$$context_A(\alpha_i) \leftarrow \beta.valence_A(\alpha_i) + (1-\beta)context_A(\alpha_{i-1}) \tag{4}$$

$$context_B(\alpha_i) \leftarrow \beta.valence_B(\alpha_i) + (1-\beta)context_B(\alpha_{i-1}) \tag{5}$$

Calculating aggregate valence levels for the characters in a story allows the interpretation framework to track their changing emotions to each other over time, at least on a gross level. Although it is highly reductive, this gross level allows performers to distil complex emotions into simple but expressive physical actions. Because they are calculated as a function of the valence of current and past actions, these levels are both summative and persistent, and thus well-suited to making decisions regarding summative and persistent physical actions in a performance. If a significant increase in positive valence for a character A is interpreted as a result of actions involving character B, then performer A can move a step closer to performer B in physical space. Conversely, a significant decrease can cause A to move a step away from B. This increase or decrease for A is given by **Eq. 6**. The same spatial/emotional calculus applies to B's perspective, as given in **Eq. 7**. In each case, a significant increase or decrease is determined to be a positive or negative change that exceeds a fixed threshold τ. In this way, the relative spatial movements of performers on stage are not explicitly indicated by the script, or directly associated with the actions in the plot, but determined by each performer's evolving interpretation of the narrative context.

$$\Delta_A(\alpha_i) \leftarrow context_A(\alpha_i) - context_A(\alpha_{i-1}) \tag{6}$$

$$\Delta_B(\alpha_i) \leftarrow context_B(\alpha_i) - context_B(\alpha_{i-1}) \tag{7}$$

The emotional valence of an action for a character, much like a character's "inertial" contextual valence, is derived from four emotionally charged scales that have been chosen to suit our system's inventory of 800 plot actions. New parallel scales can be added, or existing ones removed or replaced, if this inventory were to change significantly. Currently, one obvious omission is an arousal scale (Kensinger and Schacter, 2006), to show the degree to which an action either calms or arouses a particular role. Arousal is not a charged dimension – for one can be as aroused by hate as by love – and so it does not contribute to our calculations of valence. Nonetheless, an arousal dimension is useful for indicating the scale of an actor's response. A high-arousal action may demand a bigger and more dramatic physical response that a calming, low-arousal event. For that reason, it makes sense to add a new scale as follows:

$$\text{calmed} \leftarrow \text{A} \rightarrow \text{aroused} \quad \text{calmed} \leftarrow \text{B} \rightarrow \text{aroused}$$

A state of high-arousal can be conveyed with a sweeping, high-energy gesture. while a calm state might be conveyed with a slow movement or a slight gesture. Of course, the robot platform may not support the distinction between high- and low-energy motions. The extent to which it does, or does not, indicates the extent to which an arousal dimension is worthwhile in a story-telling context. Still, we may find that arousal is wrapped up with the question of contextual valence and how quickly the influence of context should decay. If arousal can be shown to influence the rate of decay, it would be a valuable addition to the framework whatever robot platform is used. It thus remains a topic of ongoing research in this project.

5 EVALUATION

When a performer's spatial movements and gestures are chosen on the basis of its interpretation of the plot, we deem those physical actions to be *coherent*. Conversely, when those movements and gestures at chosen at random, to create the mere appearance of embodied performance, we deem those actions *incoherent*. Clearly, the value of interpretation lies in the audience's ability to recognize coherent uses of movement and gesture. More importantly,

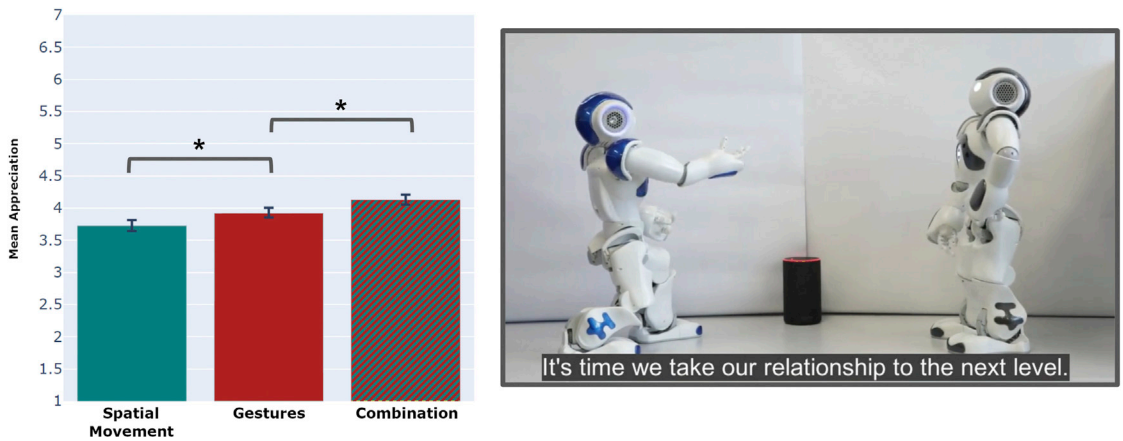


FIGURE 6 | Scenario and results using the *performance framework*. The graph shows that a combined spatial and gesture performance receives a higher average rating on a questionnaire assessing Human-Robot interaction utility measures. The star indicates a significant difference between the mean values ($p < 0.05$) and the whiskers show the standard error of the mean. The image on the right shows an example of an iconic gesture performed by one of the robots during the evaluated performance.

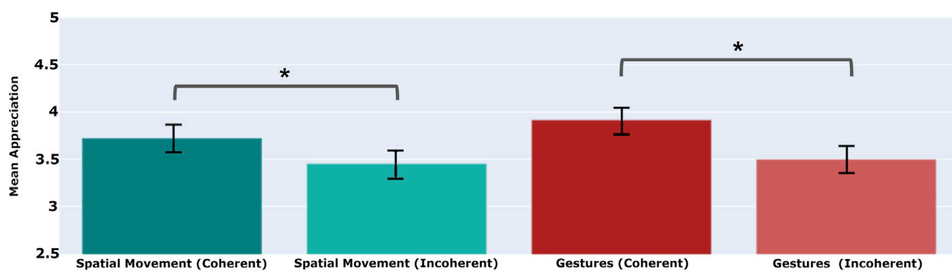


FIGURE 7 | Coherent use of spatial movement in a performance – to and fro movements dictated by a performer's interpretation of the current plot action – are appreciated more by audiences than incoherent movements that are chosen at random, or chosen contrary to that interpretation. The same can be seen for coherent vs. incoherent gestures. A star indicates a significant difference between the mean values ($p < 0.05$), while the whiskers show the standard error of the mean.

it lies in the increased appreciation that an audience will feel for coherent vs. incoherent performances. This is what we set out to evaluate here, by asking: how much do audiences appreciate spatial and gestural embodiment, and how much do they appreciate the interpretation that goes into the embodied choices that are made by our robotic performers when enacting a story?

5.1 Space and Gesture: Together and Apart

These stories are generated using *Scéalextric*, and performed by two Softbank NAO robots working within a related framework, named *Scéalability*, that choreographs their actions (Wicke and Veale, 2020b). Although the robots appear to speak to each other during the performance, the choreography is actually achieved using backstage communication *via* a blackboard architecture (Hayes-Roth, 1985). The NAO robots are bipedal, humanoid robots offering 25 degrees of freedom (Gouaillier et al., 2009). *Scéalextric* and *Scéalability* are used to generate and perform the stories that crowd-sourced judges will evaluate for spatial and gestural coherence. The test performances can combine spatial movements and gestures, or use spatial movements alone, or use gestures alone. In each case, spatial movements and gestures can be chosen coherently, on the basis of an interpretation of the plot, or incoherently. Incoherent spatial movements are chosen to be the opposite of what would be considered coherent in context; thus, if the coherent movement is to take a step forward, the incoherent alternative is to take a step backwards, and vice versa. For gestures, which have no clear opposite, the incoherent choice is a random choice amongst all available gestures.

Our first experiment has three conditions, and raters on the crowd-sourcing platform *Amazon Mechanical Turk* (*AMT*) are presented with stories reflecting one of these conditions. In the first condition, the performers use only spatial movements in the enactment, and not gestures. Those movements are always chosen to be coherent with the plot. In the second condition, the performers use only gestures, not spatial movements, where again those gestures are chosen to be coherent. In the third condition, performers use coherent spatial movements and coherent gestures in the same performance. Raters reflect their appreciation of the given performance, which they view as an online video, on a Human-Robot Interaction questionnaire. Their feedback then forms the basis of a between-subjects study. On average, and as shown in **Figure 6**, raters appear to appreciate coherent gestures more than coherent spatial movements, but appear to appreciate a coherent combination of both more than any one mode of physical expression.

For the experimental setup of three conditions, the null hypothesis (H_0) states that there are no differences in the appreciation of the performances with different types of body movements, i.e., the rating of the performance should be independent of the movement performed. Our alternative hypothesis (H_1) is that there are real differences in the appreciation of performances that use different body movements, i.e., the rating of a performance should be dependent on the type and coherence of the movements used. An analysis of variance (ANOVA) is conducted to determine whether there are any statistically significant differences between the means of the three conditions. Unless one or more of the distributions is highly skewed, or the variances are very different, the ANOVA is a reliable analytical measure. When we compare the variances of the distributions with a Levene test for homoscedasticity (Brown and Forsythe, 1974), no significant differences are found (Test-statistics = 2.207 with $p > 0.05$ to accept equal variances). Hence, we assume equal variance for all conditions.

Consequently, the analysis of variance reveals significant differences between the three conditions, with $p = 0.0019$ (Sum of squares = 38.686, F-values = 6.292). With the results of this ANOVA, we can reject the null hypothesis to argue that there is a significant dependence of the rating of the performance on the type and coherence of the movements used. We hypothesize that some movements are more appreciated than others, but since the ANOVA does not make any claims about individual differences and effects, we conducted an additional t-test. Given the significant differences between conditions of more-or-less equal variance, we applied a two-sided t-test to tease out the differences between the three conditions. Because the t-test only provides a p-value to signify in-between differences, we also calculated Cohen's d to measure an appropriate effect size for the comparison between the means of the three conditions. This t-test showed a significant difference between the spatial movement and combined movement conditions (p = 0.002 Bonferroni corrected). The spatial movement condition yielded a mean appreciation score of $\mu_{Spatial} = 3.728$ with a standard deviation of $\sigma_{Spatial} = 1.792$. The combined condition received an average appreciation rating of $\mu_{Combined} = 4.131$ with a standard deviation of $\mu_{Combined} = 1.762$. The measured effect favours the latter (Cohen's D = 0.227). Statistical tests have been conducted on the accumulated test construct (of all 14 items) and the results are visualized in **Figure 6**. More details are available in (Wicke and Veale, 2020a).

Within the Human-Robot Interaction questionnaire, each participant rates the robotic performance using an appreciation construct that comprises two parts of seven questions apiece. One part, which measures the perceived attractiveness of the performance, uses the *AttrakDiff* questionnaire of Hassenzahl et al. (2003). The other elicits quality ratings for the embodiment, e.g., as to whether the physical actions of the performers appear to be appropriate to the story. The scores in **Figure 7** represent mean average scores, which to say, mean scores for all fourteen questions averaged across all raters for the relevant conditions.

The stories generated using *Scéalextric* can be long and convoluted, with many sub-plots and secondary characters. This complexity tends to confound the analysis of embodiment choices, since it requires raters to watch long video performances. So, for the three conditions studied here, raters are shown extracts from longer performances that focus on specific events in a story that involve the kinds of movements we aim to evaluate. Three one-minute videos are shown to 40 raters for each condition (so 120 in total) on *Amazon Mechanical Turk*. Each one-minute video is an excerpt of an embodied performance with a narrative voice-over. Each rater is paid 0.40$ per video to fill out the questionnaire of 14 questions. Three additional gold-standard questions are also included, which allow us to detect disengaged raters who provide uniform or random responses. When we exclude these invalid responses, there are 32 valid responses for the Spatial Movement condition, 29 for the gesture condition and 33 for the combined movement condition, yielding a total of $N = 94$ valid responses. More details on this study can be found in (Wicke and Veale, 2020a).

The crowd-sourcing of raters on platforms such as *Amazon's Mechanical Turk (or AMT)* brings with it some clear advantages and disadvantages. AMT does not provide demographic information about its participants, so we did not seek out this information. While there are concerns about the demographic characteristics of AMT rater populations (Chandler and Shapiro, 2016), AMT can still provide a relatively diverse demography, especially if compared to other Web-based samples and to the average American campus sample (Buhrmester et al., 2016). A study of *AMT* workers by Michel et al. (2018) reports that the average age of task participants is 35.5 years (SD = 11.0), and that 58% of workers are female.

5.2 To Interpret or Not: Coherence Versus Incoherence

This first experiment concerns performances in which performers always make coherent choices. In a second experiment, we aim to show that audiences appreciate coherence more than incoherence – and thus appreciate interpretation over non-interpretation – by showing raters performances in which choices are made either coherently (using the interpretation framework) or incoherently (ignoring, or going against the interpretation framework). This second experiment creates performances that observe one of four conditions: using spatial movements alone (coherent), with no gestures; using spatial movements alone (incoherent), with no gestures; using gestures alone (coherent), with no spatial movements; and using gestures alone (incoherent), with no spatial movements. Once again, raters evaluate video performances from a given condition, and provide their ratings using the same Human-Robot Interaction questionnaire. A between-subjects study of their ratings, again collected *via Amazon Mechanical Turk* and that again incorporates gold-standard questions to weed out disengaged raters, yields the findings shown in **Figure 7**.

As before, 40 raters were recruited for each condition ($N = 40 \times 4 = 160$), and each was paid 0.40$ for completing the questionnaire after watching a 1-minute video. After filtering invalid responses, the four trials resulted in 29 valid responses for coherent gestures, 28 for incoherent gestures, 32 for coherent spatial movements and 29 for incoherent spatial movements ($N = 118$). Our findings suggest that audiences do appreciate coherent interpretation over the incoherent lack of interpretation when performers use physical actions to convey a story.

6 CONCLUSION

6.1 Frameworks for Storytelling

The relative strengths of the Performance and Interpretation frameworks underline the distinction between what is performed and how it is interpreted. Both systems are distinct, yet they must work together, because interpretation is based on performance, and the latter is shaped by what the system and its actors wish to convey. Performers must first interpret for themselves what they wish an audience to subsequently interpret from their actions. But this is hardly a novel concern. Within the theory and practice of acting, it is suggested that "Rather than playing an emotion, actors are advised to play the action and encode the emotion in the action through parameters, such as speed, intensity, shape, and direction." (El-Nasr, 2007). Human approaches to acting, such as that famously outlined in (Stanislavski, 2013), can thus inform our approach to the robotic performance of stories. Importantly, however, we must abstract away from the physical limitations and peculiar affordances of the actors themselves, or, in our case, of the specific robots that we employ. The modularity of our approach is a clear advantage in this regard.

Creativity by a producer always requires a corresponding (if perhaps lesser) creativity in the consumer if it is to be properly appreciated. In this paper we have necessarily focused on producer-side creativity, and said little about the consumer-side creativity that it necessitates in turn. This lopsided view is tenable in the short-term, for practical reasons, but it must be redressed eventually, Future work must thus address this imbalance, which is inherent to the creative equation in any performative context. Producers anticipate how consumers will react, while consumers model the intent of the producer. To adequately account for one side of the equation, we must also account for the other.

6.2 Discussion

The Softbank NAO robots that are used in our performances and crowd-sourced evaluations have none of the grace or agility exhibited by recent, bio-inspired machines, such as those of *Boston Dynamics* (Guizzo, 2019). Those robots are capable of animal-like movements and human-like poise, as recently shown in scripted robotic dances[2]. Nonetheless, we focus on a larger point here, one that is mostly independent of the robotic hardware that is used. To perform a story for an audience, performers must do more than follow a literal script to the letter. They must interpret the script, to actually fill the positions – spatially and otherwise – of the characters they are supposed to "inhabit." Interpretation requires an emotional understanding of the unfolding plot, so that actions can be chosen to coherently reflect that understanding.

To this end, our interpretation and performance frameworks employ representations and mechanisms that mediate between plot actions, character emotions, and a performer's movements and gestures on stage. Interpretation creates choice for a performer, motivating departures from the script when the scripted response seems inadequate in context. Moreover, interpretation guides performance, so that the robotic performers become part of a larger whole, in which relative position is as important as individual action.

One dimension of human emotional expression that is overlooked here is that of facial expression. We humans communicate with our looks as well as our words and gestures, as shown e.g., by the importance of non-manual features in sign language (Nguyen and Ranganath, 2012). We do not consider this dimension here because it is not the primary focus of our current work, not least because our robots lack the means to emote with their faces. Nonetheless, facial emotion is a dimension we must address in future work and in any addition to the current framework. To begin, we are presently considering the role of facial emotion and gestures in audience members as they watch a robotic performance. When viewers engage with a story, their expressions and gestures can subtly (and not so subtly) guide the interpretations of the robotic performers, perhaps warranting a comment in response, or even a dramatic plot change in mid-narrative. As outlined in (Wicke and Veale, 2021), the underlying stories can be generated as disjunctive trees rather than linear paths, and robots can elicit emotional feedback from users *via* a video feed, to determine which forks in the plot to follow.

We have defined a modular and extensible framework that can be adapted and reused by HRI researchers for different kinds of robotic performance. For example, a study on the perception of drone movements by Bevins and Duncan (2021) evaluates how participants respond to a selection of schematic flight paths. Since the identified movements are inherently schematic, e.g. Up-Down and Left-Right, we believe that our framework can help to categorize their results from a performative and an interpretative perspective. Other recent work suggests how we might extend our performance framework's taxonomy of motion types to include, as noted earlier, properties such as *naturalness*.

[2]https://www.youtube.com/watch?v=fn3KWM1kuAw

In this vein, Kitagawa et al. (2021) investigate how robots can most naturally move toward their goals, and show that common *rotate-while-move* and *rotate-then-move* strategies are inferior to their proposed set of human-inspired motion strategies.

A natural complement to the visual modality of spatial movement is sound. Three types of artificial sounds for robotic movements are explored in Robinson et al. (2021) in the context of a *Smooth Operator*. Their results indicate that robotic movements are interpreted differently, and perceived as more or less elegant, controlled or precise, when they are coupled with different sounds. These findings suggest that our interpretation framework might be applied to additional modalities, such as sound, to exploit additional channels of communication and augmented modes of meaning-making.

Of course, these modalities are usually bidirectional. The visual modality, for example, works in both directions: as the audience perceives the performers, the performers can also perceive their audience. As noted earlier, this allows audience members to use their own gestural and facial cues to communicate approval or disapproval to the actors as they enact their tales. This feedback – in the guise of a smile or a frown, a thumbs up or a thumbs down – allows performers to change tack and follow a different path through the narrative space when, as described in Wicke and Veale (2021), the underlying story contains branching points at which actors should seek explicit feedback from the audience. A video camera provides a visual feed that is analyzed for schematic gestures and facial emotions, and when such cues are present, the actors base their choice not on their own interpretation but on that of the audience.

6.3 The Nao Robot

When considering the limitations of this work, we must address our choice of robot, the Nao. Qualities that are desirable from an interpretative perspective may be undesirable from a performance perspective, and vice versa. For instance, a decision to link arousal with the energy and speed of the robot's actions must consider issues of unwanted noise (from the robot's gears) and balance (it may fall over if it reacts too dramatically). The latter also affects its use of space. As robots move closer together, to e.g., convey emotional closeness, their gestures must become more subtle, lest they accidentally strike one another in the execution of a sweeping motion. While our interpretation and performance frameworks might look different than they are had we chosen a different platform, we are confident that these modular and extensible frameworks can grow to accommodate other choices in the future, either by us or by others.

Because the Nao platform has been used in a variety of related research [e.g., Gelin et al. (2010), Pelachaud et al. (2010), Ham et al. (2011), LaViers and Egerstedt (2012), and Wicke and Veale (2018a)], this speaks well to the reproducability of our approach. Despite its limitations, the Nao currently suits our needs, not least because it has 25 degrees of freedom and the ability to move its limbs independently of each other. The robot's fixed facial expression is certainly a limitation, one that prevents us from conveying emotion with facial cues, yet this also helps us to avoid unwanted bias in our user studies. It also means that we need not worry that the robot's manual gestures will occlude its facial expressions at key points. Other limitations can be addressed in a more-or-less satisfactory fashion. For instance, the Nao cannot turn on the spot, but turning can be implemented as a composition of spatial and rotational movements. So, although the movements of our performance framework are shaped in large part by the abilities of our robots, they are not wholly determined by their limitations. A comparison with other robot platforms would undoubtedly be useful and revealing, but it is beyond the scope of this current paper.

6.4 Current Thoughts, Future Directions

The interpretation and performance frameworks support both fine-grained and gross-level insights into the unfolding narrative. For instance, we have seen that aggregate assessments of valence – for a given role of a specific action at a particular point in the plot – allow for aggregate judgments about characters and their changing feelings. These gross judgments, which reveal positive or negative shifts in a character's overall feelings, can suggest equally reductive actions for robot performers to execute on stage, such as moving closer to, or further away from, other performers playing other characters. In this way, gross interpretations support powerful spatial metaphors that are equally summative and equally persistent. We have largely focused here on the semantics of gross spatial actions, but the literature provides a formal basis for the more fine-grained forms of expression that we will pursue in future work, such as those from the domain of dance (LaViers et al., 2014; Bacula and LaViers, 2020).

But fine-grained insights are also supported by the framework, which is to say, insights based on movements along a single emotional dimension. Spatial movement to and fro, of the kind evaluated in the previous section, are reductive and general. But metaphors that allow a performer to construe on action as another in a given context, such as by construing an *insult* as an *attack*, or an act of *praise* as an act of *worship*, are more more specific. They work at the conceptual level of plot action, and do more than suggest an embodied response. Rather, they increase the range of choices available to a performer because they operate at a deeper and more specific level of interpretation.

We humans reach for a metaphor when we want to broaden our palette of expressive options, and so too can our robot performers. But metaphor it itself just one choice that leads to others. Irony is another. A performer can, for example, choose to react ironically to a script directive. Suppose character A is expected to show fealty to character B, and the story so far firmly establishes this expectation in the minds of the audience (and in the view of the interpretation framework). Irony is always a matter of critiquing a failed expectation, by acting as though it has not failed while clearly showing that it has. It is the ultimate creative choice. Suppose now that the plot calls for A to *rebel against*, or *stand up to*, or to *break with* B. When the interpretation framework compares the emotions established by previous actions with those stirred by this new action, it recognizes a rift that should, if it is large enough, influence how the performers react. The robot portraying character A might thus act out an action more in line with the expected emotions, such as bowing down to B, while speaking the dialogue associated with the current action, such as "I've had enough of you!". The bifurcation of irony, of expectation vs. reality, easily maps onto the parallel modalities of speech and physical action, so that a performer can indeed follow both branches at once.

Although we have not examined or evaluated irony here, we mention it now to show that robotic performances of complex semiotic structures, such as stories, open many avenues for an interpretative performer, at both the conceptual and the expressive levels. These choices, which include construal mechanisms such as metaphor and irony and more besides, open more choices in turn, if a performer has the wit to perceive and exploit them. As such, it is fair to say that we have barely scratched the surface of what an interpretative approach to embodied performance can yet bring to domains such as story-telling. As we go deeper, we may need to use a richer model of the gestures and motions that realize the embodiment, such as by drawing on insights and representations from the world of dance, where more nuanced actions – and more nuanced notations – necessarily hold sway.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Movement and Gesture Repository for Robots and Humans (at OSF) https://osf.io/e5bn2/?view_only=2e30ee7e715342d59c371b5d30c014e0.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the authors' university, University College Dublin, Ireland, under protocol number UCD HREC-LS, Ref. No.: LS-E-19-125-Wicke-Veale. The study received exemption from ethics approval. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Both authors contributed to conception and design of the research. The first implemented the robotic frameworks and prepared an initial draft of the manuscript. The second implemented the story-generation framework. Each author contributed equally to all sections of the manuscript, to revising the manuscript, and to preparing and submitting the current version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2021.662182/full#supplementary-material

REFERENCES

Andrist, S., Tan, X. Z., Gleicher, M., and Mutlu, B. (2014). "Conversational Gaze Aversion for Humanlike Robots," in 2014 9th ACM/IEEE International Conference on Human-Robot Interaction(HRI), Bielefeld, Germany, March, 2014 (New York, NY: Association for Computing Machinery), 25–32.

Augello, A., Infantino, I., Manfré, A., Pilato, G., and Vella, F. (2016a). Analyzing and Discussing Primary Creative Traits of a Robotic Artist. *Biol. Inspired Cogn. Archit.* 17, 22–31. doi:10.1016/j.bica.2016.07.006

Augello, A., Infantino, I., Manfrè, A., Pilato, G., Vella, F., and Chella, A. (2016b). Creation and Cognition for Humanoid Live Dancing. *Rob. Auton. Syst.* 86, 128–137. doi:10.1016/j.robot.2016.09.012

Bacula, A., and LaViers, A. (2020). Character Synthesis of Ballet Archetypes on Robots Using Laban Movement Analysis: Comparison between a Humanoid and an Aerial Robot Platform With Lay and Expert Observation. *Int. J. Soc. Robot.* 12, 1–16. doi:10.1007/s12369-020-00695-0

Bergen, B., Narayan, S., and Feldman, J. (2003). "Embodied Verbal Semantics: Evidence from an Image-Verb Matching Task," in Proceedings of the Annual Meeting of the Cognitive Science Society, Boston, MA (Austin, TX: Cognitive Science Society), 25.

Besold, T. R., Hedblom, M. M., and Kutz, O. (2017). A Narrative in Three Acts: Using Combinations of Image Schemas to Model Events. *Biol. Inspired Cogn. Archit.* 19, 10–20. doi:10.1016/j.bica.2016.11.001

Bevins, A., and Duncan, B. A. (2021). "Aerial Flight Paths for Communication: How Participants Perceive and Intend to Respond to Drone Movements," in Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, Boulder CO, March, 2021 (New York, NY: Association for Computing Machinery), 16–23.

Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms*. Hove, United Kingdom: Psychology Press.

Brand, M., and Hertzmann, A. (2000). "Style Machines," in Proceedings of the 27th annual conference on Computer graphics and interactive techniques (New York, NY: ACM Press), 183–192.

Brandt, L., and Brandt, P. A. (2005). Making Sense of a Blend: A Cognitive-Semiotic Approach to Metaphor. *Annu. Rev. Cogn. Linguist.* 3, 216–249. doi:10.1075/arcl.3.12bra

Bravo Sánchez, F. Á., González Correal, A. M., and González Guerrero, E. (2017). Interactive Drama with Robots for Teaching Non-technical Subjects. *J. Human-Robot Inter.* 6, 48–69. doi:10.5898/jhri.6.2.bravo

Bregler, C. (1997). "Learning and Recognizing Human Dynamics in Video Sequences," in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (IEEE), San Juan, PR, June 17–19, 1997 (Piscataway, NJ: IEEE), 568–574.

Bremner, P., and Leonards, U. (2016). Iconic Gestures for Robot Avatars, Recognition and Integration with Speech. *Front. Psychol.* 7, 183. doi:10.3389/fpsyg.2016.00183

Brooks, A. G., and Breazeal, C. (2006). "Working with Robots and Objects: Revisiting Deictic Reference for Achieving Spatial Common Ground," in Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, Salt Lake City, UT, March, 2006 (New York, NY: Association for Computing Machinery), 297–304.

Brown, M. B., and Forsythe, A. B. (1974). Robust Tests for the Equality of Variances. *J. Am. Stat. Assoc.* 69, 364–367. doi:10.1080/01621459.1974.10482955

Bruce, A., Knight, J., Listopad, S., Magerko, B., and Nourbakhsh, I. R. (2000). "Robot Improv: Using Drama to Create Believable Agents," in Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation Symposia Proc. (Cat. No. 00CH37065) (IEEE), San Francisco, California, April 24–28, 2000 (Piscataway, NJ; IEEE Robotics and Automation Society), 4, 4002–4008.

Bucholtz, M., and Hall, K. (2016). *Embodied Sociolinguistics. Sociolinguistics: Theoretical Debates*. Cambridge, United Kingdom: Cambridge University Press, 173–197.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2016). Amazon's Mechanical Turk: A New Source of Inexpensive, yet High-Quality Data? *Perspect. Psychol. Sci.* 6, 3–5. doi:10.1177/1745691610393980

Camurri, A., Lagerlöf, I., and Volpe, G. (2003). Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques. *Int. J. Human-Computer Stud.* 59, 213–225. doi:10.1016/s1071-5819(03)00050-8

Catala, A., Theune, M., Gijlers, H., and Heylen, D. (2017). "Storytelling as a Creative Activity in the Classroom," in Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition, Singapore, June, 2017 (New York, NY: Association for Computing Machinery), 237–242.

Chandler, J., and Shapiro, D. (2016). Conducting Clinical Research Using Crowdsourced Convenience Samples. *Annu. Rev. Clin. Psychol.* 12, 53–81. doi:10.1146/annurev-clinpsy-021815-093623

Cienki, A. (2013a). "Gesture, Space, Grammar, and Cognition," in *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives/P. Auer*. Editors M. Hilpert, A. Stukenbrock, and B. Szmrecsanyi (Berlin, Germany: Walter de Gruyter), 667–686.

Cienki, A. (2013b). Image Schemas and Mimetic Schemas in Cognitive Linguistics and Gesture Studies. *Ann. Rev. Cogn. Linguist.* 11, 417–432. doi:10.1075/rcl.11.2.13cie

Clair, A. S., Mead, R., and Matarić, M. J. (2011). "Investigating the Effects of Visual Saliency on Deictic Gesture Production by a Humanoid Robot," in 2011 RO-MAN (IEEE), Atlanta, GA, July 31–Aug 03, 2011 (Piscataway, NJ: IEEE), 210–216.

Cocks, N., Morgan, G., and Kita, S. (2011). Iconic Gesture and Speech Integration in Younger and Older Adults. *Gesture* 11, 24–39. doi:10.1075/gest.11.1.02coc

Cooney, M., and Menezes, M. (2018). Design for an Art Therapy Robot: An Explorative Review of the Theoretical Foundations for Engaging in Emotional and Creative Painting with a Robot. *Multimodal Technol. Interact.* 2, 52. doi:10.3390/mti2030052

Costa, S., Brunete, A., Bae, B.-C., and Mavridis, N. (2018). Emotional Storytelling Using Virtual and Robotic Agents. *Int. J. Hum. Robot.* 15, 1850006. doi:10.1142/s0219843618500068

Csapo, A., Gilmartin, E., Grizou, J., Han, J., Meena, R., Anastasiou, D., et al. (2012). "Multimodal Conversational Interaction with a Humanoid Robot," in 2012IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom), Kosice, Slovakia, Dec 2–5, 2012 (Piscataway, NJ: IEEE), 667–672.

Del Vecchio, D., Murray, R. M., and Perona, P. (2003). Decomposition of Human Motion into Dynamics-Based Primitives with Application to Drawing Tasks. *Automatica* 39, 2085–2098. doi:10.1016/s0005-1098(03)00250-4

El-Nasr, M. S. (2007). Interaction, Narrative, and Drama: Creating an Adaptive Interactive Narrative Using Performance Arts Theories. *Interact. Stud.* 8, 209–240. doi:10.1075/is.8.2.03eln

Fabiano, F., Pelikan, H. R., Pingen, J., Zissoldt, J., Catala, A., and Theune, M. (2017). "Designing a Co-creative Dancing Robotic Tablet," in 6th International Workshop on Computational Creativity, Concept Invention, and General Intelligence, Madrid, Spain, Dec 15, 2017.

Falomir, Z., and Plaza, E. (2020). Towards a Model of Creative Understanding: Deconstructing and Recreating Conceptual Blends Using Image Schemas and Qualitative Spatial Descriptors. *Ann. Math. Artif. Intell.* 88, 457–477. doi:10.1007/s10472-019-09619-9

Fauconnier, G., and Turner, M. (1998). Conceptual Integration Networks. *Cogn. Sci.* 22, 133–187. doi:10.1207/s15516709cog2202_1

Fauconnier, G., and Turner, M. (2008). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York, NY: Basic Books.

Fischer-Lichte, E., and Riley, J. (1997). *The Show and the Gaze of Theatre: A European Perspective*. Iowa City, IA: University of Iowa Press.

Gelin, R., d'Alessandro, C., Le, Q. A., Deroo, O., Doukhan, D., Martin, J.-C., et al. (2010). "Towards a Storytelling Humanoid Robot," in 2010 AAAI Fall Symposium Series, Arlington, Virginia, November 11–13, 2010.

Goldin-Meadow, S., and Brentari, D. (2017). Gesture, Sign, and Language: The Coming of Age of Sign Language and Gesture Studies. *Behav. Brain Sci.* 40, e46. doi:10.1017/s0140525x15001247

Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., et al. (2009). "Mechatronic Design of Nao Humanoid," in 2009 IEEE International Conference on Robotics and Automation (IEEE), Kobe, Japan, May 12-17, 2009 (Piscataway, NJ: IEEE), 769–774.

Guizzo, E. (2019). By Leaps and Bounds: An Exclusive Look at How Boston Dynamics Is Redefining Robot Agility. *IEEE Spectr.* 56, 34–39. doi:10.1109/mspec.2019.8913831

Ham, J., Bokhorst, R., Cuijpers, R., van der Pol, D., and Cabibihan, J.-J. (2011). "Making Robots Persuasive: The Influence of Combining Persuasive Strategies (Gazing and Gestures) by a Storytelling Robot on its Persuasive Power," in International Conference on Social Robotics, Amsterdam, Netherlands, November 24–25, 2011 (Berlin, Germany: Springer), 71–83.

Häring, M., Bee, N., and André, E. (2011). "Creation and Evaluation of Emotion Expression With Body Movement, Sound and Eye Color for Humanoid Robots," in 2011 RO-MAN (IEEE), Atlanta, GA, July 31–Aug 03, 2011 (Piscataway, NJ: IEEE), 204–209. doi:10.1109/roman.2011.6005263

Hassenzahl, M., Burmester, M., and Koller, F. (2003). "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität," in Mensch & computer 2003 (Springer) (Wiesbaden, Germany: Springer Vieweg Verlag), 187–196. doi:10.1007/978-3-322-80058-9_19

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron* 41, 301–307. doi:10.1016/s0896-6273(03)00838-9

Hayes-Roth, B. (1985). A Blackboard Architecture for Control. *Artif. Intell.* 26, 251–321. doi:10.1016/0004-3702(85)90063-3

Hedblom, M. M. (2020). *Image Schemas and Concept Invention: Cognitive, Logical, and Linguistic Investigations*. Basingstoke, United Kingdom: Springer Nature.

Heider, F., and Simmel, M. (1944). An Experimental Study of Apparent Behavior. *Am. J. Psychol.* 57, 243–259. doi:10.2307/1416950

Hoffman, G. (2007). Ensemble: Fluency and Embodiment for Robots Acting with Humans. PhD thesis. Cambridge (MA): Massachusetts Institute of Technology.

Hoffman, G., and Weinberg, G. (2011). Interactive Improvisation with a Robotic Marimba Player. *Auton. Robot* 31, 133–153. doi:10.1007/s10514-011-9237-0

Huang, C.-M., and Mutlu, B. (2013). Modeling and Evaluating Narrative Gestures for Humanlike Robots. *Robotics: Sci. Syst.* 9, 57–64. doi:10.15607/RSS.2013.IX.026

Johnson, M. (2013). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago, IL: University of Chicago Press.

Jokinen, K., and Wilcock, G. (2014). "Multimodal Open-Domain Conversations With the Nao Robot," in *Natural Interaction With Robots, Knowbots and Smartphones* (Berlin, Germany: Springer), 213–224.

Jouravlev, O., Zheng, D., Balewski, Z., Le Arnz Pongos, A., Levan, Z., Goldin-Meadow, S., et al. (2019). Speech-accompanying Gestures are not Processed by the Language-Processing Mechanisms. *Neuropsychologia* 132, 107132. doi:10.1016/j.neuropsychologia.2019.107132

Katevas, K., Healey, P. G., and Harris, M. T. (2014). "Robot Stand-Up: Engineering a Comic Performance," in Proceedings of the Workshop on Humanoid Robots and Creativity at the IEEE-RAS International Conference on Humanoid Robots Humanoids, Madrid, Spain, November 2014.

Kelly, S. D., Barr, D. J., Church, R. B., and Lynch, K. (1999). Offering a Hand to Pragmatic Understanding: The Role of Speech and Gesture in Comprehension and Memory. *J. Mem. Lang.* 40, 577–592. doi:10.1006/jmla.1999.2634

Kendon, A. (1980). "Gesticulation and Speech: Two Aspects of the Process of Utterance," in *The Relationship of Verbal and Nonverbal Communication* (Berlin, Germany: De Gruyter Mouton), 207.

Kensinger, E., and Schacter, D. (2006). Processing Emotional Pictures and Words: Effects of Valence and Arousal. *Cogn. Affect. Behav. Neurosci.* 6, 110. doi:10.3758/CABN.6.2.110

Kettebekov, S., and Sharma, R. (2001). "Toward Natural Gesture/speech Control of a Large Display," in IFIP International Conference on Engineering for Human-Computer Interaction (Berlin, Germany: Springer), 221–234.

Kitagawa, R., Liu, Y., and Kanda, T. (2021). "Human-inspired Motion Planning for Omni-Directional Social Robots," in Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, Boulder CO, March, 2021 (New York, NY: Association for Computing Machinery), 34–42.

Klein, S., Aeschlimann, J. F., Balsiger, D. F., Converse, S. L., Foster, M., Lao, R., et al. (1973). Automatic Novel Writing: A Status Report. University of Wisconsin-Madison Department of Computer Sciences, Tech. Rep.

Knight, H. (2011). "Eight Lessons Learned About Non-Verbal Interactions Through Robot Theater," in International Conference on Social Robotics (Berlin, Germany: Springer), 42–51.

Kuzuoka, H., Suzuki, Y., Yamashita, J., and Yamazaki, K. (2010). "Reconfiguring Spatial Formation Arrangement by Robot Body Orientation,"in 2010 5th ACM/IEEE International Conference on Human-Robot Interaction(HRI), Osaka, Japan, March 2–5, 2010 (Piscataway, NJ: IEEE), 285–292.

Ladewig, S. H. (2014). "Recurrent Gestures," in *Body–Language–Communication: An international Handbook on Multimodality in Human Interaction* (Berlin, Germany: Mouton De Gruyter), 2, 1558–1575.

Lakoff, G., and Johnson, M. (2008). *Metaphors we Live by*. Chicago, IL: University of Chicago press.

Lakoff, G. (2008). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago, IL: University of Chicago press.

LaViers, A., and Egerstedt, M. (2012). "Style Based Robotic Motion," in 2012 American Control Conference (ACC) (IEEE) (Piscataway, NJ: IEEE), 4327–4332.

LaViers, A., Teague, L., and Egerstedt, M. (2014). "Style-based Robotic Motion in Contemporary Dance Performance," in *Controls and Art* (Berlin, Germany: Springer), 205–229.

Levinson, S. C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge, United Kingdom: Cambridge University Press, Vol. 5.

Li, B., Zook, A., Davis, N., and Riedl, M. O. (2012). "Goal-driven Conceptual Blending: A Computational Approach for Creativity," in Proceedings of the 2012 International Conference on Computational Creativity, Dublin, Ireland, May 30–June 1, 2012, 3–16.

McNeill, D. (2008). *Gesture and Thought*. Chicago, IL: University of Chicago press.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago press.

McNeill, D. (1985). So You Think Gestures Are Nonverbal? *Psychol. Rev.* 92, 350. doi:10.1037/0033-295x.92.3.350

Michel, J. S., O'Neill, S. K., Hartman, P., and Lorys, A. (2018). Amazon's Mechanical Turk as a Viable Source for Organizational and Occupational Health Research. *Occup. Health Sci.* 2, 83–98. doi:10.1007/s41542-017-0009-x

Mittelberg, I. (2018). Gestures as Image Schemas and Force Gestalts: A Dynamic Systems Approach Augmented with Motion-Capture Data Analyses. *Cognitive Semiotics* 11, 394–405. doi:10.1515/cogsem-2018-0002

Mittelberg, I. (2007). "Methodology for Multimodality," in *Methods in Cognitive Linguistics* (Amsterdam, Netherlands: John Benjamins Publishing Company), 225–248.

Mittelberg, I. (2019). Peirce's Universal Categories: On Their Potential for Gesture Theory and Multimodal Analysis. *Semiotica* 2019, 193–222. doi:10.1515/sem-2018-0090

Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D., and Tessendorf, S. (2013). *Body-Language-Communication*. Berlin, Germany: Walter de Gruyter, Vol. 1.

Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., and Ishiguro, H. (2012). Conversational Gaze Mechanisms for Humanlike Robots. *ACM Trans. Interact. Intell. Syst.* 1, 1–33. doi:10.1145/2070719.2070725

Nakanishi, H., Murakami, Y., Nogami, D., and Ishiguro, H. (2008). "Minimum Movement Matters: Impact of Robot-Mounted Cameras on Social Telepresence," in Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, San Diego, CA, November, 2008 (New York, NY: Association for Computing Machinery), 303–312.

Nakauchi, Y., and Simmons, R. (2002). A Social Robot that Stands in Line. *Auton. Robots* 12, 313–324. doi:10.1023/a:1015273816637

Nguyen, T. D., and Ranganath, S. (2012). Facial Expressions in American Sign Language: Tracking and Recognition. *Pattern Recognit.* 45, 1877–1891. doi:10.1016/j.patcog.2011.10.026

Niculescu, A., van Dijk, B., Nijholt, A., Li, H., and See, S. L. (2013). Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy. *Int. J. Soc. Robot.* 5, 171–191. doi:10.1007/s12369-012-0171-x

Norris, S. (2011). Three Hierarchical Positions of Deictic Gesture in Relation to Spoken Language: A Multimodal Interaction Analysis. *Vis. Commun.* 10, 129–147. doi:10.1177/1470357211398439

Núñez, R. E., and Sweetser, E. (2006). With the Future Behind Them: Convergent Evidence From Aymara Language and Gesture in the Crosslinguistic Comparison of Spatial Construals of Time. *Cogn. Sci.* 30, 401–450. doi:10.1207/s15516709cog0000_62

Padden, C. A. (2016). *Interaction of Morphology and Syntax in American Sign Language*. London, United Kingdom: Routledge.

Peirce, C. S. (1902). "Logic as Semiotic: The Theory of Signs," in *Philosophical Writings of Peirce* (New York, NY: Dover Publications), Vol. 100.

Pelachaud, C., Gelin, R., Martin, J.-C., and Le, Q. A. (2010). "Expressive Gestures Displayed by a Humanoid Robot during a Storytelling Application," in *New Frontiers in Human-Robot Interaction (AISB)* (Leicester, GB: Society for the Study of Artificial Intelligence and Simulation of Behaviour).

Reyes, M. E., Meza, I. V., and Pineda, L. A. (2019). Robotics Facial Expression of Anger in Collaborative Human–Robot Interaction. *Int. J. Adv. Robot. Syst.* 16, 1729881418817972. doi:10.1177/1729881418817972

Ritschel, H., Aslan, I., Sedlbauer, D., and André, E. (2019). "Irony Man: Augmenting a Social Robot With the Ability to use Irony in Multimodal Communication With Humans," in Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, Montreal, QC, May, 2019 (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), 86–94.

Robinson, F. A., Velonaki, M., and Bown, O. (2021). "Smooth Operator: Tuning Robot Perception Through Artificial Movement Sound," in Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, Boulder, CO, March, 2021 (New York, NY: Association for Computing Machinery), 53–62.

Rond, J., Sanchez, A., Berger, J., and Knight, H. (2019). "Improv With Robots: Creativity, Inspiration, Co-Performance," in 2019 28th IEEE International Conference on Robot and Human Interactive Communication RO-MAN, New Delhi, India, Oct 14–18, 2019 (Piscataway, NJ: IEEE), 1–8.

Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K., and Joublin, F. (2012). Generation and Evaluation of Communicative Robot Gesture. *Int. J. Soc. Robot.* 4, 201–217. doi:10.1007/s12369-011-0124-9

Salem, M., Rohlfing, K., Kopp, S., and Joublin, F. (2011). "A Friendly Gesture: Investigating the Effect of Multimodal Robot Behavior in Human-Robot Interaction," in Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication (Piscataway, NJ: IEEE), 247–252.

Sekine, K., and Kita, S. (2017). The Listener Automatically Uses Spatial Story Representations From the Speaker's Cohesive Gestures When Processing Subsequent Sentences Without Gestures. *Acta Psychol.* 179, 89–95. doi:10.1016/j.actpsy.2017.07.009

Seo, J.-H., Yang, J.-Y., Kim, J., and Kwon, D.-S. (2013). "Autonomous Humanoid Robot Dance Generation System Based on Real-Time Music Input," in 2013 IEEE RO-MAN, Gyeongju, South Korea, Aug 26–29, 2013 (Piscataway, NJ: IEEE), 204–209.

Shamsuddin, S., Ismail, L. I., Yussof, H., Zahari, N. I., Bahari, S., Hashim, H., et al. (2011). "Humanoid Robot Nao: Review of Control and Motion Exploration," in 2011 IEEE International Conference on Control System, Computing and Engineering, Penang, Malaysia, Nov 25–27, 2011 (Piscataway, NJ: IEEE), 511–516.

Sharma, R., Yeasin, M., and Kettebekov, S. (2008). *Prosody Based Audio/visual Co-Analysis for Co-Verbal Gesture Recognition*. US Patent 7 (321), 854.

Singh, K., Davis, N., Hsiao, C.-P., Jacob, M., Patel, K., and Magerko, B. (2016). "Recognizing Actions in Motion Trajectories Using Deep Neural Networks," in Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (Vancouver, BC: PKP Publishing Services Network), Vol. 12.

Sowa, T., and Wachsmuth, I. (2009). "A Computational Model for the Representation and Processing of Shape in Coverbal Iconic Gestures," in *Spatial Language and Dialogue* (Oxford, United Kingdom: Oxford Scholarship Online). doi:10.1093/acprof:oso/9780199554201.001.0001

Stanislavski, C. (2013). *An actor Prepares (A&C Black)*. London, United Kingdom: Bloomsbury Publishing.

Straube, B., Green, A., Bromberger, B., and Kircher, T. (2011). The Differentiation of Iconic and Metaphoric Gestures: Common and Unique Integration Processes. *Hum. Brain Mapp.* 32, 520–533. doi:10.1002/hbm.21041

Sugimoto, M., Ito, T., Nguyen, T. N., and Inagaki, S. (2009). "Gentoro: A System for Supporting Children's Storytelling Using Handheld Projectors and a Robot," in Proceedings of the 8th International Conference on Interaction Design and Children, Como, Italy, June, 2009 (New York, NY: Association for Computing Machinery), 214–217.

Thörn, O., Knudsen, P., and Saffiotti, A. (2020). "Human-Robot Artistic Co-Creation: A Study in Improvised Robot Dance," in 29th IEEE International Conference on Robot and Human Interactive Communication RO-MAN), Naples, Italy, Aug 31–Sept 4, 2020 (Piscataway, NJ: IEEE), 845–850.
Veale, T., and Keane, M. T. (1992). Conceptual Scaffolding: A Spatially Founded Meaning Representation for Metaphor Comprehension. *Comput. Intell.* 8, 494–519. doi:10.1111/j.1467-8640.1992.tb00377.x
Veale, T., Wicke, P., and Mildner, T. (2019). "Duets Ex Machina: On the Performative Aspects of" Double Acts" in Computational Creativity," in Proceedings of the 10th International Conference on Computational Creativity (ICCC), Charlotte, NC, 57–64.
Vilhjálmsson, H., Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S., et al. (2007). "The Behavior Markup Language: Recent Developments and Challenges," in International Workshop on Intelligent Virtual Agents (Berlin, Germany: Springer), 99–111.
Vilk, J., and Fitter, N. T. (2020). "Comedians in Cafes Getting Data: Evaluating Timing and Adaptivity in Real-World Robot Comedy Performance-Robot Interaction," in Proceedings of the 2020 ACM/IEEE International Conference on Human, Cambridge, United Kingdom, March, 2020 (New York, NY: Association for Computing Machinery), 223–231.
Weber, K., Ritschel, H., Lingenfelser, F., and André, E. (2018). "Real-time Adaptation of a Robotic Joke Teller Based on Human Social Signals," in AAMAS '18: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems). doi:10.1145/3242969.3242976
Wicke, P., and Veale, T. (2021). "Are You Not Entertained? Computational Storytelling With Non-verbal Interaction," in Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, Boulder, CO, March, 2020 (New York, NY: Association for Computing Machinery), 200–204.
Wicke, P., and Veale, T. (2018a). "Interview With the Robot: Question-Guided Collaboration in a Storytelling System," in ICCC18: International Conference on Computational Creativity, Salamanca, Spain, June 2018, 56–63.
Wicke, P., and Veale, T. (2020a). "Show, Don't (Just) Tell: Embodiment and Spatial Metaphor in Computational Story-Telling," in Proceedings of the Eleventh International Conference on Computational Creativity, ICCC 2020, Coimbra, Portugal, Sep 7–11, 2020. Editors F. A. Cardoso, P. Machado, T. Veale, and J. M. Cunha (Portugal: Association for Computational Creativity), 268–275.
Wicke, P., and Veale, T. (2018b). "Storytelling by a Show of Hands: A Framework for Interactive Embodied Storytelling in Robotic Agents," in Proc. Of AISB'18, the Conf. on Artificial Intelligence and Simulated Behaviour, Liverpool, United Kingdom, April 4–6, 2018, 49–56.
Wicke, P., and Veale, T. (2020b). The Show Must go on: On the use of Embodiment, Space and Gesture in Computational Storytelling. *New Gener. Comput.* 38, 1–28. doi:10.1007/s00354-020-00106-y
Wicke, P., and Veale, T. (2018c). "Wheels Within Wheels: A Causal Treatment of Image Schemas in an Embodied Storytelling System," in TriCoLore (C3GI/ISD/SCORE), Bozen, Italy (CEUR Workshop Proceedings).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

HYPOTHESIS AND THEORY
published: 16 June 2021
doi: 10.3389/frobt.2021.662030

Educational Robotics and Robot Creativity: An Interdisciplinary Dialogue

Alla Gubenko[1], Christiane Kirsch[1], Jan Nicola Smilek[1], Todd Lubart[2] and Claude Houssemand[1]*

[1]Departement of Education and Social Work, Institute for Lifelong Learning and Guidance, Luxembourg, Luxembourg, [2]Université de Paris et Université Gustave Eiffel, LaPEA, Boulogne-Billancourt, France

OPEN ACCESS

Edited by:
Alessandra Sciutti, Italian Institute of Technology (IIT), Italy

Reviewed by:
Vicky Charisi, European Commission, Joint Research Center, Belgium
Maud Besancon, University of Rennes 2–Upper Brittany, France
Laurine Peter, University of Rennes 2–Upper Brittany Rennes, France, in collaboration with reviewer MB

***Correspondence:**
Claude Houssemand
claude.houssemand@uni.lu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *31 January 2021*
Accepted: *31 May 2021*
Published: *16 June 2021*

Citation:
Gubenko A, Kirsch C, Smilek JN, Lubart T and Houssemand C (2021) Educational Robotics and Robot Creativity: An Interdisciplinary Dialogue. Front. Robot. AI 8:662030. doi: 10.3389/frobt.2021.662030

There is a growing literature concerning robotics and creativity. Although some authors claim that robotics in classrooms may be a promising new tool to address the creativity crisis in school, we often face a lack of theoretical development of the concept of creativity and the mechanisms involved. In this article, we will first provide an overview of existing research using educational robotics to foster creativity. We show that in this line of work the exact mechanisms promoted by robotics activities are rarely discussed. We use a confluence model of creativity to account for the positive effect of designing and coding robots on students' creative output. We focus on the cognitive components of the process of constructing and programming robots within the context of existing models of creative cognition. We address as well the question of the role of meta-reasoning and emergent strategies in the creative process. Then, in the second part of the article, we discuss how the notion of creativity applies to robots themselves in terms of the creative processes that can be embodied in these artificial agents. Ultimately, we argue that considering how robots and humans deal with novelty and solve open-ended tasks could help us to understand better some aspects of the essence of creativity.

Keywords: creative robotics, human creativity, cognition, embodied creativity, educational robotics, human-robot collaboration, machine learning

INTRODUCTION

Enhancing the ability to generate unique and useful ideas in both humans and artificial agents is a crucial challenge for 21st-century problem solving. The ways in which humans and robots may engage in the creative process and foster the development of creative productivity is a central research question that interfaces psychology and technology. Robots have been a feature of modern culture since the early pulp fiction stories and Isaac Asimov's literary contribution. Interestingly, Robbie the Robot was one of the stars of this early period, and finally became a featured "agent" in a 1956 classic science fiction film, entitled *Forbidden Planet*. Robby the Robot, who was human-sized, possessed artificial intelligence and was a problem solver who helped humans during space missions. More recently, Robby the Robot has re-appeared, in a miniature format, as a toy that children can learn to

Abbreviations: CBR, case-based reasoning; ER, educational robotics; FC, figural creativity; MDPs, markov decision processes; RL, reinforcement learning; TCT-DP, test for creative thinking-drawing production; TLHARL, transfer learning heuristically accelerated reinforcement learning; TTCT, torrance test of creative thinking.

program. Although the idea of incorporating robots into our everyday lives might have seemed outlandish and flat-out unrealistic some decades ago, the presence of robotics has well expanded, even into classrooms.

The pedagogical motivation for connecting robots with pupils is the hypothesis that creativity may be fostered through human-machine interactive exchanges. The scientific literature highlights a number of experiments of this type which seem to produce positive effects on both children and machines. Thus, this article seeks to 1) exemplify through a synthesis of the literature what creativity-related aspects are covered by the field of educational robotics, 2) present the mechanisms underlying creativity which are potentially at work in these pedagogical situations and, thus, 3) understand better how children but also artificial agents can develop their creative expertize from physically and socially situated practices.

A SHORT OVERVIEW OF EDUCATIONAL ROBOTICS

The term "educational robotics" refers to a field of study that aims to improve student's learning experiences through the creation and implementation of activities, technologies, and artifacts related to robots (Angel-Fernandez and Vincze, 2018). In practice, these activities can involve the use of a physical robot, may that be a modular system like LEGO Mindstorms, or robots specifically constructed for the designated activities.

Such activities can be conceptualized for students from elementary to graduate levels and may include design, programming, application, or experimentation with robots. Educational robotics activities usually consist of the use of a robotics kit, with which children learn how to build and program the robots for a given task (Jung & Won, 2018). These activities can take the form of interventions, after-school activities, voluntary classes, or an entire course module focusing on robotics.

The theoretical foundations for the application of educational robots are multiple, but the constructionist educational approach has been the norm (Kafai and Resnick, 1996; Papert, 1981; Danahy et al., 2014). Robotics kits provide a modular approach regarding programming and building, often used as creativity-enhancing interventions in the school context. In working with these kits, students can exert engineering competencies and creative[1] solutions to a vast array of problems, starting from making a robot move from point A to B. Furthermore, principles such as problem-based learning and gamification are guiding the implementation of educational robotics interventions. The latter, gamification, describes the use of game elements in non-game contexts to foster motivation (Sailer et al., 2014).

The robots' humanoid appearance may foster student engagement (Zawieska et al., 2015). The characteristics of robotic devices themselves can yield interesting effects as well. In interviews with students who underwent a course including the use of robotics, Apiola et al. (2010) found that the playful aspect of robotics, partnered with the physical embodiment of learning contents, had an important role in students' engagement. An exploratory qualitative study by Nemiro et al. (2017) emphasized the role of robotics in creating an engaging classroom atmosphere.

OVERVIEW OF EXISTING INTERVENTIONS USING ROBOTICS TO FOSTER CREATIVITY

An early theoretical stance on creativity in children was developed by Vygotsky (1967), who argued that creativity would develop out of playful activities in which children engage. During these play activities, not only past experiences would be engaged, but a sort of combinatory imagination would encompass newly formed impressions stemming from new realities. Guilford (1950) asked why schools do not engage more thoroughly in the fostering of students' creative abilities.

In 1972, Papert and Solomon published "Twenty Things to Do with a Computer", in which they proposed a further integration of Information and Communication Technology into school curricula. In the article, the authors presented a robot called "Turtle", which is an early example of an educational robotics device (Papert and Solomon, 1972). This rather simplistic and non-anthropomorphic robot was directed to move around via an easy-to-learn programming language called "LOGO". Papert and Solomon described how "Turtle" could be programmed to draw pictures on the surface on which it moved via a pen that was located on the center bottom of the robot.

In the early 2000s, robotic toolkits gained an ever-growing attention in the pedagogical context (Alimisis, 2013). Wang (2001) described the use of a robotics course for engineering students, stating that LEGO robotics would be "an excellent medium for teaching design, programming and creativity" (Wang, 2001, p. 5). However, this work focused mainly on promoting engineering education content and did not include a standardized creativity measure.

Adams et al. (2010) interviewed engineering undergraduates who completed a voluntary robotics module. Among other engineering problem-solving tasks, the module involved programming a LEGO Mindstorms robot. After this module, 64% of participants stated that their creative thinking skills had improved.

Cavas et al. (2012) investigated the effect of a LEGO Mindstorms robotics course on student's scientific creativity. The sample consisted of 23 twelve-to thirteen-year-old students, attending a Turkish private school. During the course, the students were introduced to building and programming robots. The authors did not specify their measure of scientific creativity but stated that it increased in students after the program.

[1]In this article, the term "creative" refers to a response that is: adapted to the problem situation and has not been taught in class (children), adapted to the problem situation and has not been previously programmed for (robot).

Álvarez and Larrañaga (2013) examined how a robotics intervention using LEGO Mindstorms affected student's motivation and their improvement in algorithm coding abilities. Via short self-report questionnaires, the authors established an increase in the student's motivation and course interest.

Huei (2014) implemented a five-week program in which freshmore students were introduced to a programming language for coding robots. After the program, 93.25% of the 74 participants agreed or strongly agreed that the mini-project had enhanced their creativity, research and problem-solving skills (Huei, 2014). Jagust et al. (2017) presented the results of workshops for gifted elementary students using LEGO Mindstorms robotic sets. Although the authors did not psychometrically assess creativity, their qualitative analysis concluded that the children were "creatively productive" (Jagust et al., 2017).

In the context of educational robotics, the term "programming" applies also to younger pupils, considering that simple, visual programming interfaces are widely available. Using these already available or self-designed robotics kits, students are often given a specific problem to solve. Sullivan and Bers (2018) provide an example of this kind of intervention; in their study, the children were asked to program a robot to move in accordance with a given dance. During the curriculum, the researchers used Positive Technological Development checklists for observing the pupil's behavior during the intervention. Sullivan and Bers (2018) stated that the frequency of creative behavior observed during the curriculum was "relatively high" (Sullivan and Bers, 2018). Creative behavior was associated with the use of a variety of materials or with using affordances of the materials in unexpected ways.

In some studies, the effects of educational robotics on student's creativity were examined using standardized creativity measures. Alves-Oliveira (2020) investigated whether scholastic activities with robots would enhance children's creativity. Children's creativity levels were assessed in three conditions. In the first condition, children performed STEAM activities by learning how to code robots. In the second condition, children performed these activities by learning how to design robots. The third, control, condition, was comprised of children engaging in a music class. The pretest-to-posttest evolution in creativity was assessed with the Test for Creative Thinking-Drawing Production–TCT-DP (Urban and Jellen, 1996). In the TCT-DP, the examinee must finalize an unfinished drawing, and several variables, including new elements added, are evaluated. Results showed that creativity levels were boosted after each intervention. When examining the change in overall creativity scores, associated with each condition, the coding condition yielded a larger effect size than the control and the design condition. The TCT-DP assesses two creativity dimensions, namely: adaptiveness and innovativeness (Lubart et al., 2010). The effect of the design intervention on children's creativity was mainly explained by an increase in scores on the TCT-DP innovativeness dimension, which is related to unconventional ways of thinking. According to Alves-Oliveira (2020), this dimension is associated with divergent thinking.

Alves-Oliveira (2020) argued that the nature of the coding task, which involved learning via trial and error, stimulated non-conventional thinking in the children. More specifically, in the coding condition of this study, the children learned how to use "Scratch language" (Resnick et al., 2009 in Alves-Oliveira, 2020). The young participants were divided into groups of 3–4 participants. Each group was appointed to program a mail-delivery robot. The robot was directed by simple codes written by the pupils, which made the robot move from one place to another. According to Alves-Oliveira (2020), this fostered a strong effect of the coding condition on the "stimulation of non-conventional ways of thinking". The author argued that the nature of the coding task explained the larger effect size on children's "innovativeness", observed in the coding condition; the children were forced to experiment and explore during the coding tasks and learned by trial and error. Alves-Oliveira (2020) concluded that this learning via trial and error stimulated non-conventional thinking.

Eteokleous et al. (2018) conducted a study in which 32 primary school students between 5 and 12-years old participated in a 1-h non-formal robotics curriculum once per week. In order to assess the effects of the curriculum on student's creativity, the Torrance Test of Creative Thinking, TTCT (Torrance, 1974), was administered before and after the 36-week intervention. Comparisons of the creativity scores before and after the intervention indicated a significant improvement in children's creative abilities (Eteokleous et al., 2018).

Badeleh (2019) examined the effects of a robotics construction course on 120 student's creativity and physics learning. A constructivist robot learning approach was used, which means that the learning outcomes were mainly acquired through the construction and testing of a robot with the use of a prepared manual. Badeleh (2019) implemented a study design, which included an experimental and a control group. The control group received traditional physics classes. The Torrance Creativity Questionnaire (Torrance, 1974 as cited in; Badeleh, 2019), assessing the dimensions of fluidity, flexibility, innovation, and detailed explanation, was administered to both groups before and after the intervention. The results showed that the constructionist robotics training had significantly increased student's global creativity.

Hendrik et al. (2020) examined whether the use of robotics as learning tools has a positive effect on Figural Creativity (FC) in 40 elementary school students. The educational robotics intervention consisted of seven weekly lessons of 2–3 h. After the first introductory lesson, students participated in robot designing projects. To assess possible changes in FC, Hendrik et al. (2020) used the Torrance Figural Creativity Test (Torrance, 1974) before and after the intervention. Hendrik et al. (2020) defined the purposes of each lesson beforehand, and which of the four dimensions (fluency, flexibility, originality, elaboration) of the Torrance Test would be targeted each time. In one lesson, students were asked to construct an anthropomorphic robot, using LEGO Mindstorms sets. According to Hendrik et al. (2020), an important outcome of this lesson was to raise the student's attention to the fact that different types of robots (humanoid and non-humanoid) could be built with the same robotics kit. The pretest-to-posttest comparisons of global FC scores indicated that they had increased in the intervention group. Therefore, Hendrik et al. (2020) advocated the inclusion of robotics classes in school curricula.

To summarize, a substantial amount of work dedicated to Educational Robotics (ER) has been conducted. Although many studies on ER include the notion of "creativity", they refer mainly to problem-solving abilities. At times, creative abilities were exclusively assessed with self-report measures. Other studies, which relied on standardized instruments, such as the TCT-DP or the TTCT, observed increases in participant's *Innovativeness* (Alves-Oliveira, 2020), *Closure* and *Creative Strength* (Eteokleous et al., 2018). In general, studies that examined the effects of ER on creativity rarely made use of clearly defined creativity constructs, and often did not provide a detailed account of the revealed effects.

Future studies could explore the underlying cognitive aspects of ER interventions, with reference to standardized creativity measures. One line of work could investigate the specific impact of ER interventions on ideational fluency, flexibility, and originality. Another line of work could examine the differential effects of specific types of ER activities, such as differences between designing robots vs. programming robot kits for a specific task. In practice, that could result in an examination of cognitive outcomes related to either designing or programming robots. However, in order to understand the underlying cognitive processes of ER interventions, clearly defined, operationalized and transferable theoretical frameworks are necessary.

MULTIVARIATE APPROACH TO CREATIVITY–CONFLUENCE MODEL

In the multivariate approach to creativity, the confluence model (Lubart et al., 2015) considers how cognitive, conative, affective, and environmental aspects synergistically interact with the requirements of a particular field to give birth to a creative product. Cognitive aspects refer to intelligence, knowledge, and information processing abilities. Conative aspects refer to personality traits and motivation. With regards to personality, perseverance, tolerance of ambiguity, openness to new experiences, and risk taking are particularly important for creativity. The creative process does not unfold in a vacuum, however. Environment plays an important role in the translation of creative potential into a creative product.

Educational robotics provides an excellent opportunity to study how real-world creativity emerges from student's interaction with their social, physical, and cultural environment (**Figure 1**). In robotics activities, students learn to use affordances and constraints of robotic construction kits while engaging in collaborative problem solving in order to build their authentic and functional robotic device. These activities perfectly instantiate Glăveanu's definition of creativity (Glăveanu, 2013, p.76), which is "the action of an actor or group of actors, in its constant interaction with multiple audiences and the affordances of the material world, leading to the generation of new and useful artifacts".

While recognizing the role of conative factors, in this work, we will pay special attention to student's cognitive processes and strategies because we suppose that non-cognitive factors act upon cognitive ones. In the following sections, we will consider creativity as situated practice and explain the positive effect of educational robotics on student's cognitive mechanisms. However, before considering the mental process involved in robotics training, we will describe the creative process itself.

EXISTING MODELS OF CREATIVE COGNITION

One of the first models of creative thinking was proposed by Wallas (1926). His four-stage model comprised preparation (problem finding, problem analysis, and acquisition of domain skills and knowledge), incubation (putting the problem aside for a while without consciously thinking about it), illumination (a sudden burst of insight), and verification. Walla's model not only emphasized the role of meta-components such as problem definition and evaluation but also stressed the role of uncontrolled, unconscious processing in idea generation. Although the model is intuitively appealing, it has been noted that not all creative solutions arise from a spontaneous "Aha"! or "Eureka" experience. The creative idea can also be a result of deliberate problem-solving efforts (Weisberg, 1986; Finke, 1996; Dietrich, 2004). As such, a comprehensive model should give a more detailed account of cognitive operations underlying the solution-finding process. Moreover, whereas the creative process is described as linear, the real-life creative problem solving is dynamic, has a loosely structured sequence, and does not necessarily follow a linear structure (Mumford et al., 1991; Schön, 1983; Corazza and Agnoli, 2018; Lubart, 2018). Despite these drawbacks, the Walla's model (1926) has had an enormous impact on modern conceptions of the creative act and represents the first account of the creative process as involving explicit and implicit mechanisms.

Building on the model of Wallas, Amabile (1983) proposed to make a distinction between 1) the problem identification and 2) preparation stages. According to Amabile, during the former, problem definition and construction take place, whereas the latter is where reactivation of knowledge and search for task-relevant information happen. Amabile has also replaced a black-box illumination phase by 3) response generation phase and defined it as seeking and producing potential responses. She has suggested that the solution generation process represents a flexible (sometimes even random) search of possible pathways and exploring the environment's characteristics. In other words, this stage involves searching for productive heuristics, which are defined as any principle or device that provides useful shortcuts for solving novel problems. Amabile argues that the choice of strategy (a set of heuristics) is crucial as it determines the level of novelty of the final solution. This idea draws upon the information-processing model of cognition by Newell and

Simon (1972) and has received empirical support in creativity research (Spiridonov, 1997; Gilhooly et al., 2007; Nusbaum and Silvia, 2011). Newell and Simon hypothesized that people can solve unfamiliar problems because they can choose among alternative actions, anticipate the outcomes of these actions, evaluate them, and vary the approach when needed. Newell et al. (1962) called this process heuristic search through a problem space. In this view, switching between search strategies can account for the creative solution (Simon, 1986). The final step in the creative process, according to Amabile, is 4) response validation, which is similar to Walla's verification phase, and involves evaluating possible responses against factual knowledge and other criteria, along with implementing and testing the idea (Amabile, 1983; Amabile, 1996).

Concerning the incubation phase, there is evidence that some insightful ideas arise when a complex problem is temporarily set aside. Whereas some authors associated this process with the ability to abandon unproductive search strategies, i.e., "productive forgetting" (Simon, 1966; Finke, 1996), others point to the role of defocused attention (Martindale, 1999; Sarathy, 2018).

In a line of work that focuses on the component cognitive operations (Sternberg, 1986a; Sternberg, 1986b; Sternberg, 1988), or "sub-processes" that compose complex cognition, the overall creative process was examined in more detail (Lubart, 2000). The first phase of the creative process (problem definition) includes selective encoding which is responsible for updating relevant and inhibiting irrelevant information (Benedek et al., 2014) and leads to problem representation in working memory. Selective comparison is responsible for 1) recalling relevant knowledge from long-term memory, and 2) mapping the relations between new and extant knowledge (Markman and Gentner, 1993). Selective comparison allows discovering a new relationship between new and already acquired information. Finally, novel solutions during the idea generation phase arise from the combination and recombination of knowledge in working memory (Sternberg, 1988). Mumford et al. (1991) have further addressed mechanisms of knowledge combination and proposed that reasoning, analogy use, and divergent thinking account for creative solutions. Sternberg (1986b) highlights also the role of meta-components in problem finding, problem definition (and redefinition), and strategy choice. Some theorists also refer to these processes as executive functioning (Miller and Cohen, 2001).

Finke et al. (1992) developed the Geneplore model of creative cognition and distinguished between generative and exploratory phases of creative search. The idea generative phase comprises strategies such as knowledge retrieval, synthesis, and categorical reduction (see Gilhooly et al., 2007 for the description). The generative phase results in the production of *preinventive structures*—preliminary models which are characterized by novelty and ambiguity. These characteristics of preinventive structures afford numerous possibilities for the selective combination of their properties during exploratory phase. Strategies that allow further exploration of these structures are, for example, searching for potential functions, attributes or limitations, hypothesis testing, and conceptual interpretation.

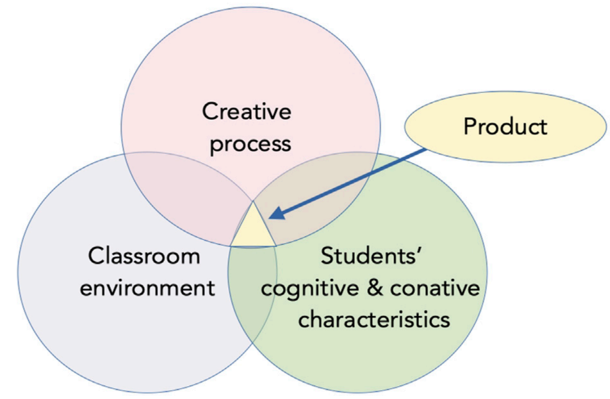


FIGURE 1 | Confluence model for educational robotics. Note: This figure is adapted from Nemiro et al. (2017).

As generation and exploration cycles repeat, the preinventive structures could be partially modified or completely replaced by the new ones.

Repetitions of Geneplore cycles and switching between generative and explorative strategies may be accompanied by changes in attentional focus. Indeed, there is evidence indicating that early stages of the creative process may involve instances of defocused attention, whereas later stages may require more focused attention (Dorfman et al., 2008; Kaufman 2011; Zabelina et al., 2016).

Martindale (1999) proposed that creative people are characterized by a better ability to shift between focused and defocused attention as a function of task demands. This claim has received empirical confirmation (Zabelina and Robinson, 2010). In terms of the Geneplore model, it means that the effective creative process may involve enhanced switching between generative and explorative strategies.

In summary, drawing on the work by Sternberg (1986a; 1986b; 1988; 2012), Amabile (1993;1996), Finke et al. (1992), Beghetto and Corazza (2019), we argue further that the creative process is a multistage dynamic process which builds on existing knowledge and is guided by a productive strategy search. This search is characterized by alternation between generative and explorative thinking. Importantly, generative and explorative cycles could unfold on two levels: a strategy could be discovered by explicitly reflecting on the task demands and previous problem-solving experience, i.e., at a meta-level, but it could also happen on the implicit level and be a result of trial and error search and exploration of associations between task, actions, and outcomes (**Figure 2**). This view is reminiscent of dual-process models (system 1, system 2) of human cognition (Crowley et al., 1997; Stanovich and West, 2000; Kahneman, 2011).

COGNITIVE COMPONENTS OF THE PROCESS OF DESIGNING AND PROGRAMMING ROBOTS

Drawing on principles of constructionism, Kolodner (2002) introduced a learning model that incorporates design and

FIGURE 2 | Two-level view of the creative process.

inquiry activities organized in two interrelated cycles: the "Investigate and Explore" cycle, where students acquire knowledge and generate ideas, and "Design/Redesign" cycle, where knowledge is applied. We can note that the model instantiates the basic principle of the Geneplore model of creative cognition (Ward et al., 1999), where the generative search alternates with explorative processes. Given the resemblance, it seems reasonable to apply existing models of creative cognition to analyze mental processes that underlie robotics activities.

The initial step in building and programming a robot is presenting the problem to be solved. For example, students are given a task to build a mobile robot and program its basic movements. This could be, for example, a creation of a robotic system that models a human heart (Cuperman and Verner, 2013), or programming a mail-delivery robot (Alves-Oliveira, 2020). A common feature of these robotic challenges is that they are poorly structured, have multiple solution paths, i.e., could be solved using different strategies, and do not have a single criterion for evaluating the solution.

From a cognitive point of view, the first step in the process of creating a robotic device is problem identification, in which a problem solver has to elaborate a problem representation. In terms of robotics, this implies analysis of the system's requirements and translation of these requirements into design specifications (Pahl and Beitz, 2007). In information processing terms, this step could be accomplished through selective encoding, i.e., selecting relevant elements of a problem and suppressing those that are not relevant for task completion (Sternberg, 1988; Benedek et al., 2014). Another important process is the retrieval of relevant information from long-term memory (Smith, 1995). Presumably, this is done via selective comparison (Sternberg, 1986a; 1986b), in which problem solver aligns existing knowledge and previous problem-solving experience with the characteristics of the new challenge (Holyoak, 1984; Mumford et al., 1991). It involves a comparison of critical elements such as goals, procedures, and constraints encountered in similar problems. In practical terms, with respect to generating ideas for a robot's design, students spend time thinking about known solutions and how they might be reused in the new task (Kolodner, 1994). This process helps learners to identify the gaps in their existing knowledge. When the problem is new and procedural and dispositional knowledge is lacking, a great deal of learning takes place (Amabile, 1983). For example, in the study of Cuperman and Verner (2013), before building a robotic model of the human heart students had to carry out investigations to learn the principle of the heartbeat mechanism. If the domain-relevant skills and knowledge are sufficient to afford a range of possible pathways to explore, students immediately start the process of building a robot after the problem has been defined.

The process of solution generation in robotics problems is often paralleled with implementation, i.e., designing the robots. As robotics problems are often ill-defined, finding possible solutions for each design specification requires a search among numerous potential alternatives within a space of possibilities (Ball et al., 1997). There is evidence that generating few ideas at this stage leads to the restriction of the search space and poor designs, as students became "fixated" on concrete solutions too early (Fricke, 1996).

The generation stage in robotics design involves mental and physical synthesis of building components and creating functional prototypes. Functional prototypes of robots that result from initial generative processes may be viewed as preinventive structures (Finke et al., 1992) that are assessed for appropriateness and other criteria and are further modified during the exploratory phase. Evaluation of the prototypes naturally leads students back to the first stages of the creative process—redefining the design specifications, as well as gathering task-relevant information (Suwa et al., 1999). This iterative process of perceiving an emerging design and making a change to it allows to learn new affordances and often leads to unexpected discoveries (Schön and Wiggins 1992; Kelly and Gero, 2014).

The process of a robot's design is followed by an iterative, trial-and-error phase of programming the robot's moves, testing, and modifying its design and software code (Nemiro et al., 2017; Alves-Oliveira, 2020; Chevalier et al., 2020). In the later cycles of the process of creation of the robotic model, students move beyond a trial-and-error method and start developing their own heuristic approach, which allows them to come up with original technical solutions (Hayes, 1978; Altshuller, 1988; Sullivan and Lin, 2012; Sullivan, 2017).

Barak and Zadok (2009) described three explorative strategies that lead learners to inventive solutions in robotic tasks. The first strategy the authors called "assigning a new function", where students find a new use for an already existing robot's movement. The second strategy involves the elimination of a component from the system. This heuristic has been extensively described in TRIZ (Altshuller, 1988). The third strategy consists of examining physical objects available in the environment and trying to apply them to solve a problem. Sullivan (2011) called this last strategy "utilizing environmental affordances". Attentional mechanisms, and more specifically, diffused attention, may be important for this strategy as it helps to notice some environmental cues leading to the generation of novel ideas (Sarathy, 2018; Zabelina, 2018).

Sullivan (2011) described the process of constructing a robotic model in terms of troubleshooting cycles and rapid prototyping rounds, in which students fluently move between 1) writing code, 2) testing the robot, 3) analyzing problems, 4) proposing changes to the model, and 5) testing the device again. The author's detailed analysis of the solution trajectory shows that each troubleshooting round includes three key stages: 1) problem identification, 2) idea generation and strategy choice, and 3) reflections on the progression of the problem-solving process. Sullivan (2011) described a case of a robotics programming activity in which the solution process consisted of 17 troubleshooting cycles and was two-fold: first, an explorative strategy was used to discover novel affordances of materials and then the problem was redefined, i.e., meta-level reasoning was applied.

To summarize, the process of building robotic models can be characterized by a constant search and movement back and forth between generative and explorative thinking (**Figure 3**). The creation of a robotic model involves using generative strategies, like memory retrieval (Sullivan, 2011), brainstorming (Nemiro et al., 2017), mental synthesis, and analogical transfer (Barak and Zadok, 2009; Cuperman and Verner, 2013), as well as explorative strategies–attribute finding, conceptual interpretation (Barak and Zadok, 2009; Chan and Schunn, 2015), and utilizing the environmental affordances (Sullivan, 2011). As our analysis suggests, the search for a solution in a robot construction process involves not only switching between generative and explorative strategies but also switching between levels of thinking at which these strategies operate. One may suppose that the practice of alternating between two different modes of cognition, generative and explorative, coupled with implicit and metacognitive processes that work in parallel, could result in better coordination between these components and promote student's cognitive flexibility. Recent instructional models for teaching creativity via educational robotics also underscore the role of generative, explorative, and meta-components (Chevalier et al., 2020; Yang et al., 2020). Another possible explanation that can account for the promotion of student's creative potential by robotics programs is that the process of engaging in collaborative construction of robotic devices leads not only to novel physical artifacts but also to the emergence of new mental tools–implicit and meta ideational strategies. Thus, engaging in physically, technologically and socially situated robotics problems could lead to the development of creative expertize in students.

This rather brief analysis does not aim to provide an exhaustive description of the process of robot building and programming. Rather, we aimed to illustrate that the solution trajectory in robotics problems could share parallels with the creative process and could be described in cognitive processing terms that are often cited in conceptions of creative cognition.

CREATIVE PROCESSES IN AUTONOMOUS ROBOTS

In previous sections, we have described creativity as a socially and materially situated practice that unfolds over time through perceiving and exploring material and technological affordances and generating novel artifacts. In addition to student's conative and cognitive factors, the confluence model of creativity emphasizes the role of the environment in translating the student's creative potential into novel and useful products. Evaluating such models of human creativity is, however, challenging in natural settings due to ethical concerns and difficulties in isolating hypothesized variables.

Modern machine learning algorithms allow roboticists to develop autonomous agents able to learn by exploring their environment. Contrary to computational creativity, research in robotics using reinforcement learning is also situated, in the sense that it uses methods applicable for embodied agents. In this regard, the robot becomes a perfect tool to study and model the emergence of creativity.

Up to this point, we have used the term "robot" in a passive form and considered it as a tool to develop human creativity. In this section, we will change our perspective to consider the robot as a testbed to implement and verify our model of the creative process. Implementing a model for physical experimentation requires specifying all internal structures and processes involved (Fong et al., 2002).

Building on the description of processes outlined in the preceding sections, we argue that to be able to simulate the creative process, autonomous agents should be able to:

1. Acquire new knowledge and learn.
2. Reactivate and reuse knowledge in a wide range of environments.
3. Select and change problem-solving strategies.
4. Use meta-reasoning to define and redefine problems, evaluate the process and artifacts.

A collection of automatic processes capable of producing behavior that would be deemed creative in humans is called a "creative system" by Wiggins. The Creative Systems Framework (Wiggins, 2006) describes the creative system in terms of a search process that goes through a conceptual space to generate artifacts. This exploratory search is coupled with a metacognitive search process that operates within all possible conceptual spaces.

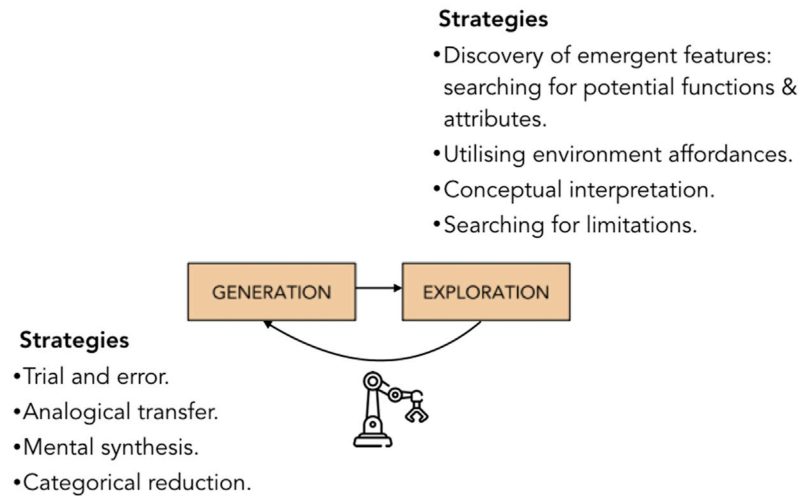


FIGURE 3 | Solution generation and exploration.

Linkola et al. (2020) attempted to apply the notion of Wiggins exploratory search to learning agents. Drawing on concepts from Markov Decision Processes (MDPs), the Creative Action Selection Framework (Linkola et al., 2020) provides a formal account of the agent's action choice based on the value, novelty, and validity of artifacts and concepts.

Several authors suggested that modern reinforcement learning algorithms based on MDPs could allow simulation of the creative process in autonomous agents (Vigorito and Barto, 2008; Schmidhuber, 2010; Colin et al., 2016). Reinforcement learning (RL) resembles the creative process as both involve interaction between a decision-making agent and its dynamic, uncertain environment, when the agent is searching for a solution to a given problem. In reinforcement learning problems, an agent explores the space of possible strategies and gets feedback based on the results of its decision making. This information is used to deduce an optimal policy (Kober et al., 2013). According to Colin et al. (2016), the agent's policy changes within hierarchical reinforcement learning algorithms resemble the change in strategies that happens during creative processes.

One of the challenges of reinforcement learning is the dilemma between exploration and exploitation (Sutton and Barto, 1998). To obtain more reward, a reinforcement learning agent must choose actions that have been effective in the past. But to discover such actions and make better action selection in the future, the robot has to try actions that it has not selected before. The creative process is also marked by the constraint between new and already existing problem-solving strategies (Collins and Koechlin, 2012) and by the necessity to build upon previous experience and knowledge in order to extend or break with them to generate novelty.

One way to address this dilemma is to introduce intrinsic motivation in RL, i.e., modifying the reward function to improve the performance of an agent (Singh et al., 2010). Whereas the traditional approach to RL is to provide reward only in case of task achievement, intrinsically motivated agents are also encouraged by "cshaping" rewards for discovering novel, surprising patterns in the environment (Ng et al., 1999). According to Schmidhuber (2010), the discovery of these novel regularities in curiosity-driven exploration would be marked by an impressive reduction in computational resources.

Recent advances in reinforcement learning are associated with the introduction of deep reinforcement learning, showcasing agents learning to play games which have long been considered as very complex for artificial agents (Mnih et al., 2015; Silver et al., 2016; Schulman et al., 2017). One of the major limitations of RL algorithms is, however, their high computational cost to learn new environments. Although RL has been successfully used to autonomously solve complex tasks, learning to solve these tasks requires large time investments. This is due to the fact that in order to converge on a good solution, RL agents require a significant number of explorative interactions with the environment.

Several approaches have been introduced to reduce reinforcement learning time; these include learning through other agent's advice in a shared environment (Saunders, 2012; Silva and Costa, 2019), and learning from human demonstrations (Argall et al., 2009; Fitzgerald et al., 2018). Another way to overcome the drawback of time-consuming exploration is to enable machine learning algorithms with the ability to transfer and reuse previously acquired knowledge across tasks using a case-based reasoning approach (CBR) (Riesbeck and Schank, 1989; Kolodner, 2014).

CBR begins with a problem representation of the situation in which the case can be used. Problem representation is compared with cases stored in a case base using specified similarity measures. If relevant cases exist, they are retrieved, adjusted, and reused in the problem at hand (Aamodt and Plaza, 1994; De Mantaras et al., 2005). Given that CBR has already been coupled with TRIZ problem-solving strategies and showed its potential to accelerate innovation design (Robles et al., 2009; Ching-Hung et al., 2019), its application to speed up RL seems promising.

Recent attempts to combine the advantages of reinforcement learning with case-based reasoning can be found in Glatt et al. (2020), Bianchi et al. (2018). Whereas Deep Case-Based Policy Inference algorithm accelerates learning by building a collection of policies and using it for a more effective exploration of a new task, the latter, Transfer Learning Heuristically Accelerated Reinforcement Learning algorithms (TLHARL), speeds up the RL process using CBR and heuristics. Bianchi et al. (2018) have shown that TLHARL improved significantly the learning rate in two domains – robot soccer and humanoid-robot stability learning.

The success of a system using CBR techniques depends on the ability of the system to retrieve, redefine, and reuse cases. To detect reasoning failures, improve the similarity assessment measure and the case adaptation mechanisms of the CBR system, meta-reasoning techniques are used. Arcos et al. (2011) have described an introspective reasoning model enabling a CBR system to learn autonomously to improve multiple facets of its reasoning process. The model performs five distinct functions: 1) monitoring the CBR process; 2) assessing the quality of proposed solutions; 3) identifying reasoning failures; 4) proposing goals; and 5) evaluating the impact of proposed improvements. Enabled with meta-reasoning, the system can identify and repair the sources of failures and thus incrementally adapt to the new problem situation.

CBR systems have their limits as well, however. Whereas they are effective when dealing with cases that bear resemblance to the task that has already been experienced by the robot, CBR systems have limited efficiency when they encounter novel problems. Parashar et al. (2018) have introduced an architecture enabling an agent to cope with novelty. The work addresses the issue raised by Sarathy and Scheutz (2018), Konidaris et al. (2018) and combines planning and reinforcement learning approaches. This combination of top-down and bottom-up approaches makes the work of Parashar et al. (2018) especially relevant for the context of creative problem solving in robotics. The authors proposed a three-layered agent architecture, with 1) object-level reasoning acts based on the information encoded from the environment; 2) deliberative reasoning, responsible for plan construction and action based on object-level information, and 3) a meta-reasoning layer responsible for problem construction and re-construction based on object-level and deliberative-level information and learning history. Meta-level reasoning also allows to control switching between object-level and deliberative strategies.

In this section, we have outlined the techniques that could be a possible starting point for modeling the creative process in artificial systems. A tentative model of system architecture is shown in **Figure 4**. A combination of these or similar techniques (Augello et al., 2018; Edmonds et al., 2020; Goel et al., 2020) might result in a hybrid approach for design agents capable of addressing novelty and handling MacGyver-type problems using affordances (Sarathy and Scheutz, 2018).

DISCUSSION

We began with the observation that whereas numerous studies have shown a positive effect of constructing and programming robots on creativity, little attention has been paid to the mechanisms that can account for this effect. Educational robotics has been considered as an inherently creative activity. To address this gap, we have examined the process of designing and programming robots with respect to existing models of creative cognition. Our analysis resulted in a description of the creative process as a multistage process, which builds on existing knowledge and involves trial-and-error, generative, explorative, and metacognitive components. Next, we reviewed some recent techniques enabling robots to simulate the creative process and proposed that a combination of reinforcement learning, case-based reasoning, and meta-reasoning methods has the potential to design robots that can address novelty and solve MacGyver-type problems.

Many questions remain, however. First, as the confluence model (Lubart et al., 2015) specifies, a combination of cognitive mechanisms is a necessary condition for the creative product to appear. Conative and environmental aspects must also join to engage creative work. And yet, what is even more striking, our current understanding of human creativity is far from complete, as psychologists still do not know precisely how these multiple factors interactively work together to influence creative production. For example, what is the optimal level of a person's intrinsic motivation and tolerance to ambiguity to achieve a creative outcome? Does intrinsic motivation enhance the use of certain strategies? How do contextual variables, such as resources or an uncooperative environment, modify the creative process? Is there a threshold for the various creativity predictors, under which creativity cannot arise? Can creativity occur if one cognitive or conative feature is completely missing?

In the case of robotics, even though certain cognitive processes have been emulated, it is still not clear how robots construct problem representations, what is the nature of these representations, or whether robots can autonomously find problems to solve. Regarding the non-cognitive aspects of Lubart et al.'s confluence model (2015), the question arises as to which extent robots can be designed to incorporate conative aspects.

In the light of conceiving robots that should act as social agents, their potential "personality" moves into the spotlight. If the genetic contribution to personality is lower than to cognition (Loehlin and Nichols, 2012), it should theoretically be easier to program robots that develop a certain "personality", and this is what some researchers have tried to do (Goetz and Kiesler, 2002; Lee et al., 2006; Woods et al., 2007; Tapus et al., 2008), notably regarding the introversion/extraversion trait (Goetz and Kiesler, 2002; Lee et al., 2006; Tapus et al., 2008). The important question is to which extent robots can imitate the major creativity-related traits, including perseverance, tolerance of ambiguity, openness to new experiences, and risk-taking (Lubart et al., 2015). Regarding openness to new experiences, which is viewed as the most relevant personality trait for creativity (McCrae, 1987; Feist, 1998; Feist, 1999), no direct attempts have been realized to program an "open-minded" robot. Agnoli et al. (2015) found that attentional processing of apparently irrelevant information (irrelevance processing) acts as a moderator between openness and creative performance. It is

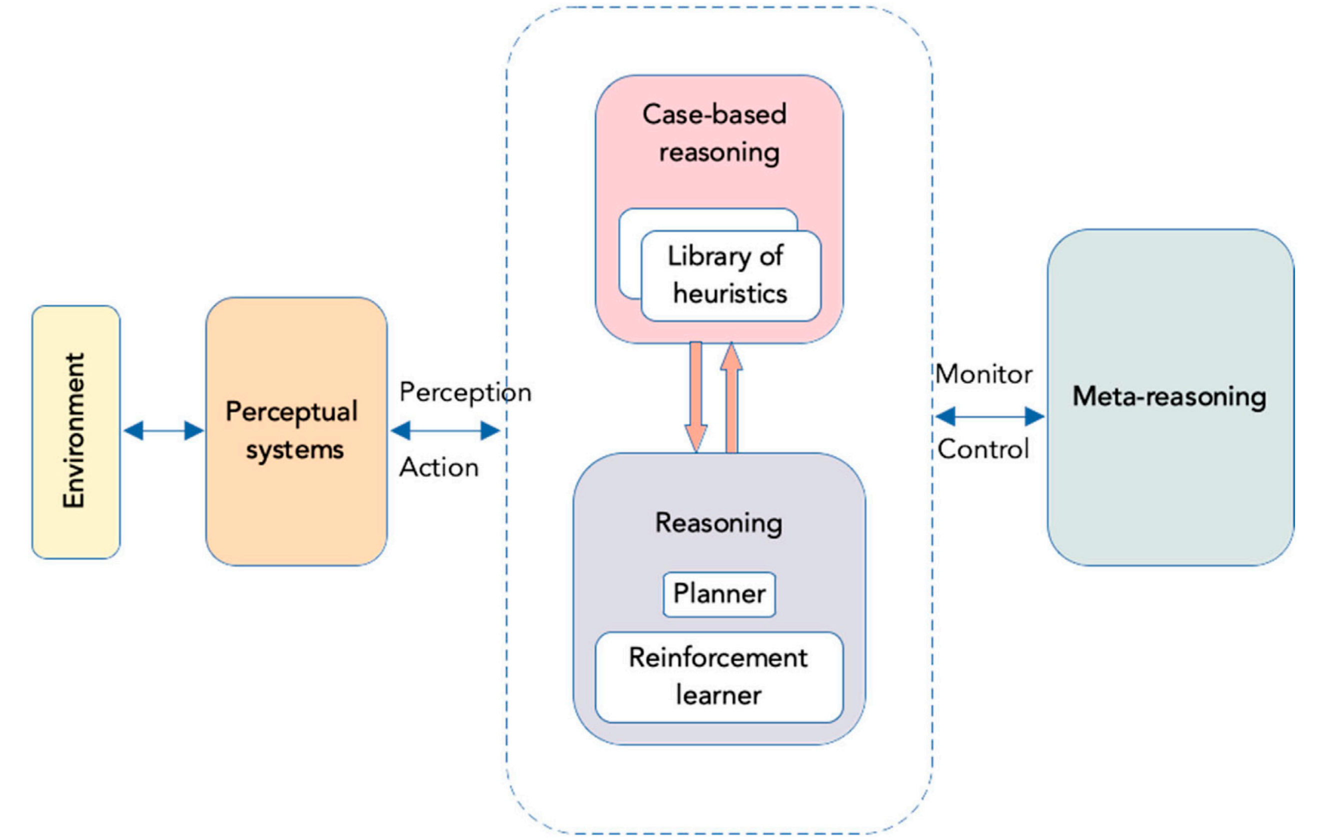


FIGURE 4 | System architecture.

imaginable that robots could be programmed for irrelevance processing and, as such, embody a certain "openness".

With respect to tolerance of ambiguity, creative performance is favored by encouraging people not to be satisfied by hasty, partial, or non-optimal solutions to complex problems (Lubart et al., 2015). Re-interpreted as a metacognitive skill, ambiguity tolerance refers to the "ability to cope with increasing sensitization to novel features of a phenomenon in order to redefine prior conceptual interpretations, contingent on trust and motivation" (Lakhana, 2012, p. III). When defined in this way, it is imaginable that robots could be programmed to display ambiguity tolerance.

As far as motivation is concerned, most attention has focused on intrinsic motivation as a positive condition for creative engagement and achievement in humans (Collins and Amabile, 1999). As described in the previous section, there are currently attempts to create intrinsically motivated robots using the reinforcement learning approach, especially regarding their intrinsically motivated open-ended learning (Schmidhuber, 2010; Santucci et al., 2020). The research is also marked with some encouraging attempts (Parisi and Petrosino, 2010; Kashani et al., 2012; Daglarli, 2020) to simulate robot's emotional states.

When it comes to the environmental aspects fostering creative performance, as we have mentioned in the previous section, there are already robots that cooperate and transfer knowledge (Silva and Costa, 2019). Projects like the Curious Whispers (Saunders et al., 2010), which study the potential of artificial society's evolution within a human physical, social, and cultural environment, are being investigated.

The possibility of comparing humans and robots in terms of creativity has traditionally focused on the productions of both, looking at whether humans and robots may produce similar or different creative work. Questions concerning the relative originality or productivity of humans and computers are raised. In contrast, our focus has been process-oriented. Do humans, who engage in a robot construction project, involve specific types of cognition that foster the development of creativity? Do robots, which instantiate artificial intelligence algorithms, engage in creative processing as humans do spontaneously? A robot may best be compared with a human baby who is learning and making discoveries by exploring the environment. As Smith and Gasser (2005), p.13 argued, "starting as a baby grounded in a physical, social, and linguistic world is crucial to the development of the flexible and inventive intelligence that characterizes humankind." We suggest that full-fledged creativity is in a robot's "zone of proximal development" (Vygotsky, 1967): what a robot cannot reach alone, it may reach with the help of a human teacher. As we have seen, robots, even in their simplest form, could also aid humans in their creative endeavors. Hence, humans and robots could fruitfully complement one another in the elaboration of creative outcomes.

CONCLUSION

In this work, we have described the creative process in information and cognitive processing terms, suggesting that computer science and cognitive psychology have had a mutual impact on each other. This influence has led to the development of a common language among psychologists and computer science engineers. As our analysis suggests, creativity research in psychology has accumulated a large set of empirical data and theoretical knowledge on human creativity, which can be useful for both an analysis of the benefits of robot design and programming for students to develop their own creativity, as well as the design of artificial agents, robots, who are themselves capable of being creative. After providing models of human creativity for machine design, psychology could gain new insights from the implementation and verification of these models in embodied agents. Interdisciplinary dialogue and collaboration between psychologists and roboticists could contribute toward better understanding of creativity and the future development of both creative humans and creative robots.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Aamodt, A., and Plaza, E. (1994). Case-based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Commun.* 7, 39–59. doi:10.3233/AIC-1994-7104

Adams, J., Kaczmarczyk, S., Picton, P., and Demian, P. (2010). Problem Solving and Creativity in Engineering: Conclusions of a Three Year Project Involving Reusable Learning Objects and Robots. *Eng. Educ.* 5 (2), 4–17. doi:10.11120/ened.2010.05020004

Agnoli, S., Franchin, L., Rubaltelli, E., and Corazza, G. E. (2015). An Eye-Tracking Analysis of Irrelevance Processing as Moderator of Openness and Creative Performance. *Creativity Res. J.* 27 (2), 125–132. doi:10.1080/10400419.2015.1030304

Alimisis, D. (2013). Educational Robotics: Open Questions and New Challenges. *Themes Sci. Tech. Educ.* 6 (1), 63–71 .

Altshuller, G. S. (1988). *Creativity as an Exact Science*. New York: Gordon & Breach.

Álvarez, A., and Larrañaga, M. (2013). *Using LEGO Mindstorms to Engage Students on Algorithm Design* [Conference Paper]. IEEE Frontiers in Education Conference. Oklahoma City, OK.. doi:10.1109/FIE.2013.6685052

Alves-Oliveira, P. (2020). *Boosting Children's Creativity through Creative Interactions with Social Robots*. (Lisbon, Portugal: University Institute of Lisbon). [unpublished doctoral dissertation].

Amabile, T. M. (1996). *Creativity in Context*. Boulder, CO: Westview Press.

Amabile, T. M. (1993). Motivational Synergy: toward New Conceptualizations of Intrinsic and Extrinsic Motivation in the Workplace. *Hum. Resource Manag. Rev.* 3 (3), 185–201. doi:10.1016/1053-4822(93)90012-S

Amabile, T. M. (1983). The Social Psychology of Creativity: a Componential Conceptualization. *J. Personal. Soc. Psychol.* 45 (2), 357–376. doi:10.1037/0022-3514.45.2.357

Apiola, M., Lattu, M., and Pasanen, T. A. (2010). "*Creativity And Intrinsic Motivation In Computer Science Education: Experimenting With Robots* [Conference Paper]," in Annual Conference on Innovation and Technology (Vilnius, Lithuania: Computer Science Education). doi:10.1145/1822090.1822147

Arcos, J. L., Mülâyim, O., and Leake, D. B. (2011). "Using Introspective Reasoning to Improve CBR System Performance," in *Metareasoning : Thinking about Thinking*. Editors M. T. Cox and Raja (Cambridge, MA: MIT Press), 21–28.

Argall, B. D., Chernova, S., Veloso, M., and Browning, B. (2009). A Survey of Robot Learning from Demonstration. *Robotics Autonomous Syst.* 57 (5), 469–483. doi:10.1016/j.robot.2008.10.024

Augello, A., Infantino, I., Maniscalco, U., Pilato, G., Rizzo, R., and Vella, F. (2018). *Robotic Intelligence and Computational Creativity. Encyclopedia with Semantic Computing and Robotic Intelligence*. World Scientific Publishing. 1850011. doi:10.1142/S2529737618500119

Badeleh, A. (2019). The Effects of Robotics Training on Students' Creativity and Learning in Physics, *Educ. Inf. Technol.* 26, 1353–1365. doi:10.1007/s10639-019-09972-6

Ball, L. J., St.B.T. Evans, J., Dennis, I., and Ormerod, T. C. (1997). Problem-solving Strategies and Expertise in Engineering Design. *Thinking & Reasoning* 3 (4), 247–270. doi:10.1080/135467897394284

Barak, M., and Zadok, Y. (2009). Robotics Projects and Learning Concepts in Science, Technology and Problem Solving. *Int. J. Technol. Des. Educ.* 19 (3), 289–307. doi:10.1007/s10798-007-9043-3

Beghetto, R. A., and Corazza, G. E. (2019). *Dynamic Perspectives on Creativity*. New York: Springer International.

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., and Neubauer, A. C. (2014). Intelligence, Creativity, and Cognitive Control: The Common and Differential Involvement of Executive Functions in Intelligence and Creativity. *Intelligence* 46, 73–83. doi:10.1016/j.intell.2014.05.007

Bianchi, R. A. C., Santos, P. E., da Silva, I. J., Celiberto, L. A., and Lopez de Mantaras, R. (2018). Heuristically Accelerated Reinforcement Learning by Means of Case-Based Reasoning and Transfer Learning. *J. Intell. Robot Syst.* 91, 301–312. doi:10.1007/s10846-017-0731-2

Cavas, B., Kesercioglu, T., Holbrook, J., Rannikmae, M., Ozdogru, E., and Gokler, F. (2012). "The Effects of Robotics Club on the Students' Performance on Science Process and Scientific Creativity Skills and Perceptions on Robots, Human and Society," in Proceedings of 3rd International Workshop teaching robotics, teaching with robotics integrating robotics in school curriculum. Editors M. Moro and D. Alimisis (Riva del Garda, Italy: TRTWR 2012).

Chan, J., and Schunn, C. (2015). The Impact of Analogies on Creative Concept Generation: Lessons from anIn VivoStudy in Engineering Design. *Cogn. Sci.* 39, 126–155. doi:10.1111/cogs.12127

Chevalier, M., Giang, C., Piatti, A., and Mondada, F. (2020). Fostering Computational Thinking through Educational Robotics: a Model for Creative Computational Problem Solving. *Int. J. STEM Ed.* 7 (39), 1–18. doi:10.1186/s40594-020-00238-z

Colin, T. R., Belpaeme, T., Cangelosi, A., and Hemion, N. (2016). Hierarchical Reinforcement Learning as Creative Problem Solving. *Robotics Autonomous Syst.* 86, 196–206. doi:10.1016/j.robot.2016.08.021

Collins, A., and Koechlin, E. (2012). Reasoning, Learning, and Creativity: Frontal Lobe Function and Human Decision-Making. *Plos Biol.* 10 (3), e1001293. doi:10.1371/journal.pbio.1001293

Collins, M. A., and Amabile, T. M. (1999). “Motivation and Creativity,” in *Handbook of Creativity*. Editor R. J. Sternberg (New York: Cambridge University Press), 297–312.

Corazza, G. E., and Agnoli, S. (2018). “The Creative Process in Science and Engineering,” in *The Creative Process. Palgrave Studies in Creativity and Culture*. Editor T. Lubart (London: Palgrave Macmillan), 155–180. doi:10.1057/978-1-137-50563-7_6

Cortes Robles, G., Negny, S., and Le Lann, J. M. (2009). Case-based Reasoning and TRIZ: A Coupling for Innovative conception in Chemical Engineering. *Chem. Eng. Process. Process Intensification* 48 (1), 239–249. doi:10.1016/j.cep.2008.03.016

Crowley, K., Shrager, J., and Siegler, R. S. (1997). Strategy Discovery as a Competitive Negotiation between Metacognitive and Associative Mechanisms. *Develop. Rev.* 17, 462–489. doi:10.1006/drev.1997.0442

Cuperman, D., and Verner, I. M. (2013). Learning through Creating Robotic Models of Biological Systems. *Int. J. Technol. Des. Educ.* 23, 849–866. doi:10.1007/s10798-013-9235-y

Daglarli, E. (2020). Computational Modeling of Prefrontal Cortex for Meta-Cognition of a Humanoid Robot. *IEEE Access* 8, 98491–98507. doi:10.1109/ACCESS.2020.2998396

Danahy, E., Wang, E., Brockman, J., Carberry, A., Shapiro, B., and Rogers, C. B. (2014). LEGO-based Robotics in Higher Education: 15 Years of Student Creativity. *Int. J. Adv. Robotic Syst.* 11, 27. doi:10.5772/58249

Dietrich, A. (2004). The Cognitive Neuroscience of Creativity. *Psychon. Bull. Rev.* 11, 1011–1026. doi:10.3758/bf03196731

Dorfman, L., Martindale, C., Gassimova, V., and Vartanian, O. (2008). Creativity and Speed of Information Processing: a Double Dissociation Involving Elementary versus Inhibitory Cognitive Tasks. *Personal. Individual Differences* 44 (6), 1382–1390. doi:10.1016/j.paid.2007.12.006

Edmonds, M., Ma, X., Qi, S., Zhu, Y., Lu, H., and Zhu, S.-C. (2020). Theory-based Causal Transfer: Integrating Instance-Level Induction and Abstract-Level Structure Learning [Conference Paper]. (Vancouver, Canada: AAAI Conference on Artificial Intelligence).

Eteokleous, N., Nisiforou, E., Christodoulou, C., Liu, L., and Gibson, D. (2018). “Fostering Children’s Creative Thinking: A pioneer Educational Robotics Curriculum,” in *Research Highlights in Technology and Teachers Education*. Editors L. Liu and D. C. Gibson (Waynesville, NC, 89–98.

Feist, G. J. (1999). “Affect in Artistic and Scientific Creativity,” in *Affect, Creative Experience and Psychological Adjustment*. Editor S. W. Russ (Philadelphia: Taylor & Francis), 9–108.

Feist, G. J. (1998). A Meta-Analysis of Personality in Scientific and Artistic Creativity. *Pers Soc. Psychol. Rev.* 2 (4), 290–309. doi:10.1207/s15327957pspr0204_5

Finke, R. A. (1996). Imagery, Creativity, and Emergent Structure. *Conscious. Cogn.* 5, 381–393. doi:10.1006/ccog.1996.0024

Finke, R. A., Ward, T. B., and Smith, S. M. (1992). *Creative Cognition: Theory, Research, and Applications*. Cambridge: The MIT Press.

Fitzgerald, T., Goel, A., and Thomaz, A. (2018). Human-guided Object Mapping for Task Transfer. *J. Hum.-Robot Interact.* 7 (2), 1–24. doi:10.1145/3277905

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2002). A Survey of Socially Interactive Robots. *Robotics Autonomous Syst.* 42 (3-4), 143–166. doi:10.1016/S0921-8890(02)00372-X

Fricke, G. (1996). Successful Individual Approaches in Engineering Design. *Res. Eng. Des.* 8, 151–165. doi:10.1007/BF01608350

Gilhooly, K. J., Fioratou, E., Anthony, S. H., and Wynn, V. (2007). Divergent Thinking: Strategies and Executive Involvement in Generating Novel Uses for Familiar Objects. *Br. J. Psychol.* 98, 611–625. doi:10.1111/j.2044-8295.2007.tb00467.x

Glatt, R., Da Silva, F. L., da Costa BianchiCosta, R. A., and Costa, A. H. R. (2020). DECAF: Deep Case-Based Policy Inference for Knowledge Transfer in Reinforcement Learning. *Expert Syst. Appl.* 156, 113420. doi:10.1016/j.eswa.2020.113420

Glăveanu, V. (2013). Rewriting the Language of Creativity: The Five A’s Framework. *Rev. Gen. Psychol.* 17, 69–81. doi:10.1037/a0033646

Goel, A. K., Fitzgerald, T., and Parashar, P. (2020). “Analogy and Metareasoning: Cognitive Strategies for Robot Learning,” in *Human-Machine Shared Contexts*. Editors W. Lawless, R. Mittu, and D. Sofge (Cambridge, MA: Academic Press), 23–44. doi:10.1016/b978-0-12-820543-3.00002-x

Goetz, J., and Kiesler, S. (2002). “Cooperation With a Robotic Assistant [Conference Paper],” in CHI’02 Extended Abstracts on Human Factors in Computing Systems (Minneapolis, MN: Association for Computing Machinery). doi:10.1145/506443.506492

Guilford, J. P. (1950). Creativity. *Am. Psychol.* 5, 444–454. doi:10.1037/h0063487

Hayes, J. R. (1978). *Cognitive Psychology Thinking and Creating*. Homewood: Dorsey Press.

Hendrik, B., Ali, N. M., and Nayan, N. M. (2020). Robotic Technology for Figural Creativity Enhancement: Case Study on Elementary School. *Int. J. Adv. Comput. Sci. Appl.* 11 (1), 536–543. doi:10.14569/ijacsa.2020.0110166

Holyoak, K. J. (1984). “Mental Models in Problem Solving,” in *Tutorials in Learning and Memory*. Editors J. R. Anderson and K. M. Kosslyn (New York: Freeman)), 193–218.

Huei, Y. C. (2014). “*Benefits And Introduction To Python Programming For Freshmore Students Using Inexpensive Robots* [Conference Paper],” in IEEE International Conference on Teaching (Wellington, New Zealand: Assessment and Learning for Engineering).

Jagust, T., Cvetkovic-Lay, J., Krzic, A. S., and Sersic, D. (2017). “Using Robotics To Foster Creativity In Early Gifted Education [Conference Paper],” in International Conference on Robotics and Education RiE, Sofia, Bulgaria.

Jung, S., and Won, E.-s. (2018). Systematic Review of Research Trends in Robotics Education for Young Children. *Sustainability* 10, 905. doi:10.3390/su10040905

Kafai, Y. B., and Resnick, M. (1996). *Constructionism in Practice: Designing, Thinking, and Learning in a Digital World*. Mahwah, NJ: Lawrence Erlbaum.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York City, NY: Macmillan.

Kashani, M. M. R., Jangjou, M., Khaefinejad, N., and Laleh, T. (2012). “Adventurous Robots Equipped with Basic Emotions,” in IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (New Orleans, LA: IEEE). doi:10.1109/CogSIMA.2012.6188362

Kaufman, S. B. (2011). “Intelligence and the Cognitive Unconscious,” in *The Cambridge Handbook of Intelligence*. Editors R. J. Sternberg and S. B. Kaufman (Cambridge, UK: Cambridge University Press), 442–467.

Kelly, N., and Gero, J. S. (2014). Interpretation in Design: Modelling How the Situation Changes during Design Activity. *Res. Eng. Des.* 25 (2), 109–124. doi:10.1007/s00163-013-0168-y

Kober, J., Bagnell, J. A., and Peters, J. (2013). Reinforcement Learning in Robotics: a Survey. *Int. J. Robotics Res.* 32 (11), 1238–1274. doi:10.1177/0278364913495721

Kolodner, J. L. (2014). *Case-Based Reasoning*. San Mateo, CA: Morgan Kaufmann.

Kolodner, J. L. (2002). Learning by Design™: Iterations of Design Challenges for Better Learning of Science Skills. *Cogn. Stud. Bull. Jpn. Cogn. Sci. Soc.* 9 (3), 338–350. doi:10.11225/jcss.9.338

Kolodner, J. L. (1994). Understanding Creativity: a Case-Based Approach. *Lecture Notes Comput. Sci.* 837, 1–20. doi:10.1007/3-540-58330-0_73

Konidaris, G., Kaelbling, L. P., and Lozano-Perez, T. (2018). From Skills to Symbols: Learning Symbolic Representations for Abstract High-Level Planning. *Jair* 61, 215–289. doi:10.1613/jair.5575

Lakhana, A. (2012). *Tolerance of Ambiguity in Educational Technology: A Review of Two Social Science Concepts*. [dissertation]. Montreal, Canada: Concordia University.

Lee, C.-H., Chen, C.-H., Li, F., and Shie, A.-J. (2020). Customized and Knowledge-Centric Service Design Model Integrating Case-Based Reasoning and TRIZ. *Expert Syst. Appl.* 143, 113062. doi:10.1016/j.eswa.2019.113062

Lee, K. M., Peng, W., Jin, S.-A., and Yan, C. (2006). Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human-Robot Interaction. *J. Commun.* 56 (4), 754–772. doi:10.1111/j.1460-2466.2006.00318.x

Linkola, S., Guckelsberger, C., and Kantosalo, A. (2020). “*Action Selection In the Creative Systems Framework* [Conference Paper],” in Eleventh International Conference on Computational Creativity, Coimbra, Portugal. Editors F. Amílcar Cardoso, P. Machado, T. Veale, and J. Miguel Cunha.

Loehlin, J. C., and Nichols, R. C. (2012). *Heredity, Environment, and Personality: A Study of 850 Sets of Twins*. Austin, TX: University of Texas Press.

Lopez De Mantaras, R., McSherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., et al. (2005). Retrieval, Reuse, Revision and Retention in Case-Based Reasoning. *Knowledge Eng. Rev.* 20 (3), 215–240. doi:10.1017/S0269888906000646

Lubart, T., Mouchiroud, C., Tordjman, S., and Zenasni, F. (2015). *Psychologie de la Créativité*. Paris: Armand Colin.

Lubart, T., Pacteau, C., Jacquet, A.-Y., and Caroff, X. (2010). Children's Creative Potential: an Empirical Study of Measurement Issues. *Learn. Individual Differences* 20 (4), 388–392. doi:10.1016/j.lindif.2010.02.006

Lubart, T. (2018). *The Creative Process: Perspectives from Multiple Domains*. London: Palgrave Macmillan.

Markman, A. B., and Gentner, D. (1993). Structural Alignment during Similarity Comparisons. *Cogn. Psychol.* 25, 431–467. doi:10.1006/cogp.1993.1011

Martindale, C. (1999). "Biological Bases of Creativity," in *Handbook of Creativity*. Editor R. J. Sternberg (Cambridge: Cambridge University Press), 137–152.

McCrae, R. R. (1987). Creativity, Divergent Thinking, and Openness to Experience. *J. Personal. Soc. Psychol.* 52 (6), 1258–1265. doi:10.1037/0022-3514.52.6.1258

Miller, E. K., and Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annu. Rev. Neurosci.* 24 (1), 167–202. doi:10.1146/annurev.neuro.24.1.167

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level Control through Deep Reinforcement Learning. *Nature* 518 (7540), 529–533. doi:10.1038/nature14236

Mumford, M. D., Mobley, M. I., Reiter-Palmon, R., Uhlman, C. E., and Doares, L. M. (1991). Process Analytic Models of Creative Capacities. *Creativity Res. J.* 4, 91–122. doi:10.1080/10400419109534380

Nemiro, J., Larriva, C., and Jawaharlal, M. (2017). Developing Creative Behavior in Elementary School Students with Robotics. *J. Creat. Behav.* 51 (1), 70–90. doi:10.1002/jocb.87

Newell, A., Shaw, J. C., and Simon, H. A. (1962). "The Processes of Creative Thinking," in *Contemporary Approaches to Creative Thinking: A Symposium Held at the University of Colorado*. Editors H. E. Gruber and M. Wertheimer (New York: Atherton Press), 63–119. doi:10.1037/13117-003

Newell, A., and Simon, H. (1972). *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.

Ng, A. Y., Harada, D., and Russell, S. (1999). *Policy Invariance under Reward Transformations: Theory and Application to Reward Shaping*. Burlington, MA: Morgan Kaufmann.

Nusbaum, E. C., and Silvia, P. J. (2011). Are Intelligence and Creativity Really So different?☆Fluid Intelligence, Executive Processes, and Strategy Use in Divergent Thinking. *Intelligence* 39 (1), 36–45. doi:10.1016/j.intell.2010.11.002

Pahl, G., and Beitz, W. (2007). *Engineering Design: A Systematic Approach*. Berlin: Springer.

Papert, S. (1981). *Mindstorms: Children, Computers, and Powerful Ideas*. UK: Harvester Press.

Papert, S., and Solomon, C. (1972). Twenty Things to Do with a Computer. *Educ. Tech.* 12 (4), 9–18 .

Parashar, P., Goel, A. K., Sheneman, B., and Christensen, H. I. (2018). Towards Life-Long Adaptive Agents: Using Metareasoning for Combining Knowledge-Based Planning with Situated Learning. *Knowledge Eng. Rev.* 33, e24. doi:10.1017/s0269888918000279

Parisi, D., and Petrosino, G. (2010). Robots that Have Emotions. *Adaptive Behav.* 18 (6), 453–469. doi:10.1177/1059712310388528

Riesbeck, C., and Schank, R. (1989). *Inside Case-Based Reasoning*. Hillsdale, NJ: Lawrence Erlbaum.

Sailer, M., Hense, J., Mandl, J., and Klevers, M. (2014). Psychological Perspectives on Motivation through Gamification. *Interaction Des. Architecture J.* 19, 28–37 .

Santucci, V. G., Oudeyer, P.-Y., Barto, A., and Baldassarre, G. (2020). Editorial: Intrinsically Motivated Open-Ended Learning in Autonomous Robots. *Front. Neurorobot.* 13, 115. doi:10.3389/fnbot.2019.00115

Sarathy, V. (2018). Real World Problem-Solving. *Front. Hum. Neurosci.* 12, 1–14. doi:10.3389/fnhum.2018.00261

Sarathy, V., and Scheutz, M. (2018). The MacGyver Test: A Framework for Evaluating Machine Resourcefulness and Creative Problem Solving. *arXiv* 1704.08350 [Preprint]. Available at: https://arxiv.org/abs/1704.08350 (Accessed January 15, 2021).

Saunders, R., Gemeinboeck, P., Lombard, A., Bourke, D., and Kocabali, B. (2010). Curious Whispers: An Embodied Artificial Creative System [Conference Paper]. Lisbon, Portugal: International Conference on Computational Creativity.

Saunders, R. (2012). Towards Autonomous Creative Systems: a Computational Approach. *Cogn. Comput.* 4 (3), 216–225. doi:10.1007/s12559-012-9131-x

Schmidhuber, J. (2010). Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990-2010). *IEEE Trans. Auton. Ment. Dev.* 2 (3), 230–247. doi:10.1109/TAMD.2010.2056368

Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*, Vol. 5126. New York: Basic Books.

Schön, D. A., and Wiggins, G. (1992). Kinds of Seeing and Their Functions in Designing. *Des. Stud.* 13 (2), 135–156. doi:10.1016/0142-694x(92)90268-f

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). *Proximal Policy Optimization Algorithms*. [preprint] Available at: arXiv: 1707.06347 (Accesesd January 15, 2021).

Silva, F. L. D., and Costa, A. H. R. (2019). A Survey on Transfer Learning for Multiagent Reinforcement Learning Systems. *jair* 64, 645–703. doi:10.1613/jair.1.11396

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., et al. (2016). Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* 529, 7587484–7587489. doi:10.1038/nature16961

Simon, H. (1966). "Scientific Discovery and the Psychology of Problem Solving," in *Mind and Cosmos: Essays in Contemporary Science and Philosophy*. Editor R. G. Colodny (Pittsburgh, PA: University of Pittsburgh Press).

Simon, H. A. (1986). The Information Processing Explanation of Gestalt Phenomena. *Comput. Hum. Behav.* 2 (4), 241–255. doi:10.1016/0747-5632(86)90006-3

Singh, S., Lewis, R. L., Barto, A. G., and Sorg, J. (2010). Intrinsically Motivated Reinforcement Learning: an Evolutionary Perspective. *IEEE Trans. Auton. Ment. Dev.* 2 (2), 70–82. doi:10.1109/TAMD.2010.2051031

Smith, L., and Gasser, M. (2005). The Development of Embodied Cognition: Six Lessons from Babies. *Artif. Life* 11, 13–29. doi:10.1162/1064546053278973

Smith, S. M. (1995). "Fixation, Incubation, and Insight in Memory and Creative Thinking," in *The Creative Cognition Approach*. Editors S. M. Smith, T. M. Ward, and R. A. Finke (Cambridge, MA: MIT Press), 135–156.

Spiridonov, V. F. (1997). The Role of Heuristic Devices in the Development of Processes in Resolving a Creative Task. *J. Russ. East Eur. Psychol.* 35 (2), 66–85. doi:10.2753/RPO1061-0405350266

Stanovich, K. E., and West, R. F. (2000). Individual Differences in Reasoning: Implications for the Rationality Debate?. *Behav. Brain Sci.* 23 (5), 645–665. doi:10.1017/s0140525x00003435

Sternberg, R. J. (1988). "A Three-Facet Model of Creativity," in *The Nature of Creativity: Contemporary Psychological Perspectives*. Editor R. J. Sternberg (Cambridge: Cambridge University Press), 125–147.

Sternberg, R. J. (1986a). "Synopsis of a Triarchic Theory of Human Intelligence," in *Intelligence and Cognition*. Editors S. H. Irvine and S. E. Newstead (Dorrecht, Germany: Nijhoff), 161–221.

Sternberg, R. J. (2012). "The Triarchic Theory of Successful Intelligence," in *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. Editors D. P. Flanagan and P. L. Harrison (New York: The Guilford Press), 156–177.

Sternberg, R. J. (1986b). Toward a Unified Theory of Human Reasoning. *Intelligence* 10, 281–314. doi:10.1016/0160-2896(86)90001-2

Sullivan, A., and Bers, M. U. (2018). Dancing Robots: Integrating Art, Music, and Robotics in Singapore's Early Childhood Centers. *Int. J. Technol. Des. Educ.* 28 (2), 325–346. doi:10.1007/s10798-017-9397-0

Sullivan, F. R., and Lin, X. D. (2012). The Ideal Science Student Survey: Exploring the Relationship of Students' Perceptions to Their Problem Solving Activity in a Robotics Context. *J. Interactive Learn. Res.* 23 (3), 273–308 .

Sullivan, F. R. (2011). Serious and Playful Inquiry: Epistemological Aspects of Collaborative Creativity. *Educ. Tech. Soc.* 14 (1), 55–65.

Sullivan, F. R. (2017). "The Creative Nature of Robotics Activity: Design and Problem Solving," in *Robotics in STEM Education*. Editor M. S. Khine (Cham): Springer International Publishing: Springer), 213–230. doi:10.1007/978-3-319-57786-9_9

Sutton, R., and Barto, A. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Suwa, M., Gero, J. S., and Purcell, T. (1999). "Unexpected Discoveries and S-Inventions of Design Requirements: A Key to Creative Designs," in *Computational Models of Creative Design IV*. Editors J. S. Gero and M-L. Maher (Sydney: Key Centre of Design Computing and Cognition, University of Sidney)), 297–320.

Tapus, A., Ţăpuş, C., and Matarić, M. J. (2008). User-robot Personality Matching and Assistive Robot Behavior Adaptation for post-stroke Rehabilitation Therapy. *Intel Serv. Robotics* 1 (2), 169–183. doi:10.1007/s11370-008-0017-4

Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking - Norms-Technical ManualResearch Edition*. Princeton, NJ: Personnel Press.

Urban, K. K., and Jellen, H. G. (1996). *Test for Creative Thinking - Drawing Production (TCTDP)*. Lisse, Netherlands: Swets and Zeitlinger.

Vigorito, C. M., and Barto, A. G. (2008). "Hierarchical Representations of Behavior for Efficient Creative Search," in AAAI Spring Symposium: Creative Intelligent Systems (Palo Alto, CA: AAAI), 135–141.

Vygotsky, L. S. (1967). Play and its Role in the Mental Development of the Child. *Soviet Psychol.* 5, 6–18. doi:10.2753/rpo1061-040505036

Wallas, G. (1926). *The Art of Thought*. London: J. Cape.

Wang, E. (2001). Teaching Freshmen Design, Creativity and Programming with Legos and Labview, Proceedings of the 31st Annualal Frontiers in Education Conference. 3. Reno, NV: Impact on Engineering and Science Education, F3G–F11. doi:10.1109/FIE.2001.963943

Ward, T. B., Smith, S. M., and Finke, R. A. (1999). "Creative Cognition," in *Handbook of Creativity*. Editor R. J. Sternberg (Cambridge: Cambridge University Press), 189–212.

Weisberg, R. W. (1986). *Creativity: Genius and Other Myths*. New York: Freeman.

Wiggins, G. A. (2006). A Preliminary Framework for Description, Analysis and Comparison of Creative Systems. *Knowledge-Based Syst.* 19, 449–458. doi:10.1016/j.knosys.2006.04.009

Woods, S., Dautenhahn, K., Kaouri, C., Boekhorst, R. t., Koay, K. L., and Walters, M. L. (2007). Are Robots like People? *Int. Studies.* 8 (2), 281–305. doi:10.1075/is.8.2.06woo

Yang, Y., Long, Y., Sun, D., Aalst, J., and Cheng, S. (2020). Fostering Students' Creativity via Educational Robotics: An Investigation of Teachers' Pedagogical Practices Based on Teacher Interviews. *Br. J. Educ. Technol.* 51, 1826–1842. doi:10.1111/bjet.12985

Zabelina, D. L. (2018). "Attention and Creativity," in *The Cambridge Handbook of the Neuroscience of Creativity*. Editors R. E. Jung and O. Vartanian (Cambridge, UK: Cambridge University Press), 161–179. doi:10.1017/9781316556238.010

Zabelina, D. L., and Robinson, M. D. (2010). Creativity as Flexible Cognitive Control. *Psychol. Aesthetics, Creativity, Arts* 4, 136–143. doi:10.1037/a0017379

Zabelina, D., Saporta, A., and Beeman, M. (2016). Flexible or Leaky Attention in Creative People? Distinct Patterns of Attention for Different Types of Creative Thinking. *Mem. Cogn.* 44, 488–498. doi:10.3758/s13421-015-0569-4

Zawieska, K., Duffy, B. R., Zieliński, C., Kaliczyńska, M., and SpringerLink (2015). "The Social Construction of Creativity in Educational Robotics," in *Advances in Intelligent Systems and Computing 351*. Editors R. Szewczyk, C. Zieliński, and M. Kaliczyńska (Switzerland: Springer International Publishing), 329–338. doi:10.1007/978-3-319-15847-1_32

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MB declared a past co-authorship with one of the authors TL to the handling editor.

BRIEF RESEARCH REPORT
published: 25 June 2021
doi: 10.3389/frobt.2021.657291

Brainstorming With a Social Robot Facilitator: Better Than Human Facilitation Due to Reduced Evaluation Apprehension?

*Julia Geerts, Jan de Wit and Alwin de Rooij**

Department of Communication and Cognition, Tilburg Center for Cognition and Communication, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands

OPEN ACCESS

Edited by:
Patricia Alves-Oliveira, University of Washington, United States

Reviewed by:
Natalia Calvo-Barajas, Uppsala University, Sweden
Tesca Fitzgerald, Carnegie Mellon University, United States
Claude Houssemand, University of Luxembourg, Luxembourg
Alla Gubenko, University of Luxembourg, Luxembourg, in collaboration with reviewer CH

****Correspondence:***
Alwin de Rooij
AlwindeRooij@tilburguniversity.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *22 January 2021*
Accepted: *14 May 2021*
Published: *25 June 2021*

Citation:
Geerts J, de Wit J and de Rooij A (2021) Brainstorming With a Social Robot Facilitator: Better Than Human Facilitation Due to Reduced Evaluation Apprehension?. Front. Robot. AI 8:657291. doi: 10.3389/frobt.2021.657291

Brainstorming is a creative technique used to support productivity and creativity during the idea generation phase of an innovation process. In professional practice, a facilitator structures, regulates, and motivates those behaviors of participants that help maintain productivity and creativity during a brainstorm. Emerging technologies, such as social robots, are being developed to support or even automate the facilitator's role. However, little is known about whether and how brainstorming with a social robot influences productivity. To take a first look, we conducted a between-subjects experiment (N = 54) that explored 1) whether brainstorming with a Wizard-of-Oz operated robot facilitator, compared to with a human facilitator, influences productivity; and 2) whether any effects on productivity might be explained by the robot's negative effects on social anxiety and evaluation apprehension. The results showed no evidence for an effect of brainstorming with a teleoperated robot facilitator, compared to brainstorming directly with a human facilitator, on productivity. Although the results did suggest that overall, social anxiety caused evaluation apprehension, and evaluation apprehension negatively affected productivity, there was no effect of brainstorming with a robot facilitator on this relationship. Herewith, the present study contributes to an emerging body of work on the efficacy and mechanisms of the facilitation of creative work by social robots.

Keywords: social robot, brainstorming, facilitator, creativity, social anxiety, evaluation apprehension

INTRODUCTION

Originally developed by Osborn (1957), the *brainstorming* technique motivates people to generate and express as many outrageous ideas as they can, while refraining from criticizing each other's ideas. In this way, they can build upon each other's ideas freely, under the assumption that quantity will ultimately lead to creativity. The role of a facilitator is to structure, regulate, and motivate those behaviors that enable participants in a brainstorm to maintain productivity and creativity throughout (Isaksen et al., 2010). For example, by enforcing brainstorm rules when participants deviate from these. However, facilitation requires advanced knowledge and skill about creative thinking that is hard to come by. Emerging technologies, such as co-creative agents and specifically social robots, are therefore increasingly looked at as an alternative to professional human facilitation (Davis et al., 2015; Frich et al., 2019). This research program is further emboldened by experimental findings that suggest that generating ideas with a social robot facilitator can enhance productivity and creativity, when compared to facilitation delivered *via* other technologies (Kahn et al., 2016; Ali et al., 2021; Alves-Oliveira et al., 2020). Although social robots are generally defined as being (semi) autonomous (Bartneck and Forlizzi 2004), in the present work we have used teleoperation to explore the potential future in which social robots would be able to autonomously facilitate brainstorming. Therefore, in the present study "social" mainly refers to the humanlike appearance and behavior of the robot as perceived by others, rather than its social intelligence. Surprisingly little is known about how working with a social robot facilitator compares to working with a human facilitator, and what the mechanisms may be that underlie its potentially advantageous effects on productivity and creativity. The present study takes a first look at how brainstorming with a social robot facilitator compares to brainstorming with a human facilitator.

Compared to virtual co-creative agents, *brainstorming with a social robot facilitator* shows great potential because these embodied machines can be designed to perceive and understand the world around them, and to communicate with humans using natural language (Fong et al., 2003). Thus, they can deliver facilitation *via* known and readily understandable communication channels, and *in situ* (Zawieska, 2014). Recent findings support that doing brainstorming and other creative work with a social robot facilitator might be advantageous over using other technologies. Alves-Oliveira et al. (2020), for example, showed that using the social robot YOLO as a character in a storytelling task, led children to generate more original ideas when YOLO actively facilitated creative thinking than when YOLO was turned off. In addition, Ali et al. (2021) showed that facilitating figural creativity by engaging and managing turn-taking in a drawing completion task by means of the social robot Jibo, increased productivity, flexibility, and originality scores of children's drawings, compared to facilitation by an iPad application. Furthermore, Kahn et al. (2016) found that facilitation by a (teleoperated) social robot led adult participants to generate more creative expressions while designing a Zen rock garden, than when facilitation was delivered *via* a PowerPoint presentation. The authors of the present paper, however, propose that understanding the true efficacy of brainstorming with a social robot facilitator also requires comparison with a human facilitator, rather than with another technology.

To explore this open scientific and applied problem, the following research question will be answered:

> "Does brainstorming with a social robot facilitator, compared to brainstorming with a human facilitator, influence productivity?"

Previous research on brainstorming in groups suggests that social interactions with other people may cause productivity losses (Sawyer 2011). Specifically, past experimental work by Camacho and Paulus (1995) showed how people that have a stronger, compared to a weaker, disposition to experience a fear of being watched or judged by others produced fewer ideas when they brainstormed with others, compared to when they brainstormed alone. Such a disposition, or *trait social anxiety*, is thought to increase the chance that people experience this anxiety in transient emotional form, *state social anxiety*, while interacting with another human being (Spielberger 1966). In turn, the social anxiety experienced may cause *evaluation apprehension* (Leary 1983; Bordia, Irmer, and Abusah 2006), where people during a brainstorm or other creative task do not express all of their ideas because they fear the social consequences of sharing these ideas (Diehl and Stroebe 1987; Warr and O'Neill 2005). Experiencing social anxiety would thus result in a productivity loss during brainstorming due to its effects on evaluation apprehension, while the likelihood that this occurs is moderated by the disposition to experience social anxiety.

Interestingly, there is also evidence that suggests that social robots can help mitigate social anxiety. A recent study by Nomura et al. (2020) showed that when anticipating collaboration, people with a stronger, compared to a weaker, disposition to experience social anxiety were more likely to prefer collaborating with a social robot than with a human being. Speculatively, this may be because some social robots tend to be perceived as non-judgmental and patient (Breazeal 2011), or because of the perception that social robots do not possess the same agency as human beings, but are rather considered as being somewhere in between inanimate toys and animate social beings (Scassellati et al., 2012). This unique relationship between human and social robot might lead people to engage in social interactions with these machines, with a decreased chance of experiencing the feeling that what they say or do is being evaluated or judged in any way by the robot. Though a mere conjecture, this previous work suggests that brainstorming with a social robot facilitator, compared to brainstorming with a human facilitator, might increase productivity because it prevents triggering a psychological mechanism where social anxiety causes evaluation apprehension to occur, with productivity loss as a consequence.

Based on these conjectures, the following working hypothesis will be explored:

> "Brainstorming with a social robot facilitator, compared to brainstorming with a human facilitator, increases productivity due to its effects on the relationship between state social anxiety and evaluation apprehension, which is moderated by trait social anxiety."

MATERIALS AND METHODS

To explore the research question and working hypothesis an experiment was conducted with a between-subjects design, where participants were asked to brainstorm with either a social robot, teleoperated by a professional human facilitator, or directly with the human facilitator.

Participants

Fifty-four people participated in the experiment (M_{age} = 23.21, SD_{age} = 3.24, $Range_{age}$ = [18, 35], 34 females, 20 males). The participants were recruited *via* the researchers' own network and

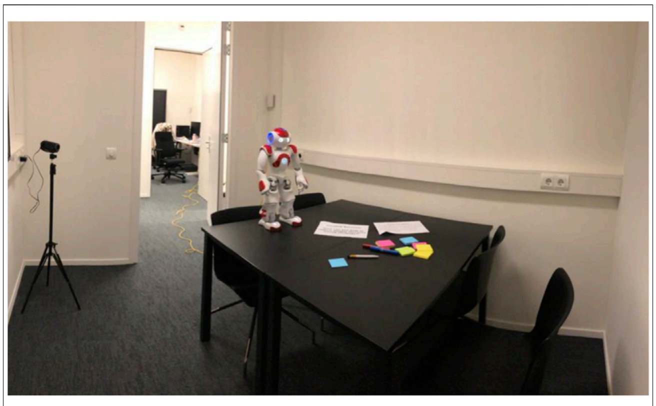


FIGURE 1 | Setup of the robot facilitator condition.

the human subjects pool of Tilburg University. The participants were predominantly Dutch ($N = 29$). Only a few participants that brainstormed with the robot facilitator had seen ($N = 7$) or collaborated ($N = 5$) with the social robot before in another situation. The participants possessed an acceptable to good level of knowledge about the brainstorming task topic ($M = 3.35$, $SD = 0.96$), and experienced an acceptable to good ability to think creatively during the facilitated brainstorm sessions ($M = 3.56$, $SD = 0.74$). Participants that were recruited through the human subject pool received study credits. The study was approved by the TSHD Research Ethics and Data Management Committee, Tilburg University.

Materials and Measures

The protocol, source code for the robot interaction, and measurement instruments are all available in the supplementary files.[1]

Brainstorming Task

The participants were asked to brainstorm ideas using Osborn's, (1957) now classical brainstorm rules for the problem: *"How can you help to reduce mental illness among students?"*. This topic was chosen for its sensitiveness and actuality among the participants (RIVM 2018). The former increases the chance that evaluation apprehension occurs (Diehl and Stroebe 1987; Pinsonneault et al., 1999). There were no criteria of what constituted an idea: this could be an initial thought, or a concrete solution (e.g., mindfulness app). All ideas were written down by the participant on Post-Its, which were color coded to indicate whether the idea originated from the participant or the facilitator. The brainstorm task took 15 min.

Robot vs. Human Facilitator

Participants were randomly assigned to brainstorm with a social robot facilitator ($N = 27$; coded: 0) or a human facilitator ($N = 27$; coded: 1). For the social robot facilitator, the Wizard-of-Oz method was used where the participants sat face-to-face with a social robot (SoftBank Robotics NAO v5) that was invisibly controlled from another room by the same professional facilitator that was present in the human facilitator condition. The Wizard-of-Oz method is used in related work as well (Kahn et al., 2016), and allowed us to maximize consistency between the two conditions. To enable robot facilitation, the Choregraphe software (Pot et al., 2009) was used to remotely send pre-defined and custom responses to the participant while a camera was used to monitor the brainstorm (**Figure 1**). The responses were vocalized to the participants through the robot's text-to-speech capabilities. The robot was "breathing" (swaying its arms and legs slightly) to simulate life-likeness, but did not use any other forms of non-verbal communication. In both conditions, the facilitator used the same response protocol. This protocol was developed to strike a balance between the rich role that facilitators play in a brainstorm, while maintaining the believability of the robot and the human as a facilitator. This entailed pre-defining short general purpose responses that covered instructions needed to structure the different phases of the brainstorm (e.g., mentioning Osborn's brainstorm rules), and process-regulating (e.g., *"Do you know another way to solve the problem?"*) and motivating messages (e.g., *"I like that idea as well!"*) needed to keep a brainstorm going. When participants were stuck or too fixated on a line of thinking, the facilitator deviated from using only pre-defined messages and relied on their experience to provide the participant with an idea to keep the brainstorm going. This unscripted assistance was provided in both conditions, and the number of facilitator-proposed ideas was counted to control for variation between participants (see *Assessment of Facilitator Intervention*). Five participants suspected or were unsure whether the social robot was controlled by a human being, but only when explicitly asked after the brainstorm and not during the brainstorm. Although an influence therefore cannot be ruled out, it is likely to be small. Thus, their data was included in the analyses to prevent an imbalanced distribution across the experimental conditions.

Assessment of Trait and State Social Anxiety

Trait and state social anxiety were both assessed using a 13-item five-point Likert scale (1 = strongly disagree, 5 = strongly agree) from the Social Interaction Anxiety Scale (Heimberg et al., 1992). Seven items were removed from the original 20-item scale because they did not apply to both trait and state anxiety. Two items were reverse coded. To assess trait anxiety the original items were administered. Cronbach alpha suggested good internal consistency, $\alpha = 0.821$. State anxiety was assessed with rephrased questions that fit the experience of social anxiety during the brainstorming task. For example, the item *"I find myself worrying that I don't know what to say in social situations"* was rephrased as *"I found myself worrying that I wouldn't know what to say in the session"*. Here, Cronbach alpha suggested minimally acceptable internal consistency, $\alpha = 0.679$. The means for the trait and state anxiety items were used in the analysis.

[1]https://osf.io/g5bhy/?view_only=6606a7779fed43188914d803bc053407.

TABLE 1 | Results principle component analysis of the evaluation apprehension questionnaire.

Items	Components evaluation apprehension		
	No room for expression	Criticism on ideas	Fear of evaluation
As collaboration partners, we listened to each other's ideas (r)	0.606	0.485	0.022
As collaboration partners, we gave each other's ideas fair consideration (r)	0.799	0.119	0.149
I was at ease during the idea generation session (r)	0.479	−0.672	−0.100
The collaboration partner was very critical in their reaction to other ideas	−0.225	0.659	0.096
I would not want my name attached to some of the ideas	0.737	−0.071	0.057
I kept thinking that the collaboration partner would criticize my ideas	−0.047	−0.111	0.970
I did not express all of my ideas because I did not want the collaboration partner to think I was weird or crazy	0.404	0.317	−0.146

Data are factor loadings for the items contained in the evaluation apprehension questionnaire. Items one to three were reverse coded (r).

Assessment of Evaluation Apprehension

Evaluation apprehension was assessed with a seven-item five-point Likert scale developed by Bolin and Neuman (2006) (1 = strongly disagree, 5 = strongly agree). The original items were reformulated to better fit the dyadic nature of the present study. For example, items such as *"As a group, we listened to . . . "* were reformulated into *"As collaboration partners, we listened to . . . "*. The first three items were reverse coded. Although previous work suggested good consistency of the scale, the Cronbach alpha in the present study was not acceptable, $\alpha = 0.181$. To check whether one or more unwieldly items may be responsible, Cronbach alphas were calculated while excluding items from the scale. This to no avail. Therefore, principle component analysis (oblique rotation) was used to explore whether the scale measured different factors (Field 2013). The results showed three factors with an eigenvalue over 1.00 that together explained 60.82% of the variance. Sampling adequacy was acceptable, $KMO = 0.60$. Inspection of the items suggested that these three factors could be interpreted as measures of "no room for expression," "criticism on ideas," and "fear of evaluation." The three factors were used in the analysis. The items and factor loadings are presented in **Table 1**.

Assessment of Productivity

To measure the participants' productivity the number of ideas they produced during the brainstorm was counted (Diehl and Stroebe 1987; Paulus and Yang 2000). This is in line with common instructions used during brainstorming in professional practice where there is an initial focus on producing many ideas (quantity), without criticizing or otherwise evaluating generated ideas (quality) (Paulus and Yang 2000). Only non-redundant ideas, written down on Post-Its by each participant, were counted.

Assessment of Facilitator Intervention

Because the facilitator intervenes at times to keep the brainstorm going by generating an idea, the number of ideas introduced by the facilitator was also counted. The number of ideas introduced may confound the tested relationships between state anxiety, evaluation apprehension, and productivity. If this is the case, these will be included as a covariate in the statistical analysis.

Demographics and Task-Relevant Sample Characteristics

Participants filled in basic demographic information (age, gender, and nationality) and were asked to "*. . . indicate your level of expertise about the topic of the brainstorm*" and "*. . . rate your level of creativity during the idea generation session*" on a five-point Likert scale (1 = very poor, 5 = very good). As knowledge is at the basis of creativity (Abraham 2018), and good facilitation entails ensuring that people feel they are creative (Isaksen et al., 2010), these are reported as relevant sample characteristics. These are reported in *Participants*.

Procedure

The study was conducted at the Media Design Lab of Tilburg University. There, participants were seated in a room at a table (**Figure 1**) and read the study information, COVID-19 protocols, task instructions, and signed informed consent. Information that could reveal the use of the Wizard-of-Oz method and the true purpose of the experiment was not yet shared. After this, the participants filled in the trait social anxiety questionnaire. Then, they engaged in the brainstorm task with either the robot or the human facilitator. After the brainstorm, the participants filled in the questionnaires used to assess state social anxiety, evaluation apprehension, and their demographics. Finally, they were fully debriefed and thanked for taking part in the experiment. After they left, the researcher recorded the number of ideas generated by the facilitator and by the participant.

RESULTS

To explore whether brainstorming with a social robot facilitator, compared to brainstorming with a human facilitator, influences productivity, an independent-samples t-test was conducted with facilitator type (robot facilitator code = 0; human facilitator code = 1) as the independent variable and productivity as the dependent variable. See **Table 2** for an overview of the descriptive statistics and correlations.

The results showed no significant difference between brainstorming with a social robot facilitator ($M = 11.33$, $SD = 2.81$) and a human facilitator ($M = 12.37$, $SD = 3.51$), for

TABLE 2 | Means, standard deviations (between parentheses), and Pearson correlations (two-tailed).

Variable	Robot facilitator	Human facilitator	Correlations						
			1	2	3	4	5	6	7
1. Productivity	11.33 (2.81)	12.37 (3.51)	–						
2. State anxiety	1.99 (0.46)	1.86 (0.44)	–0.077	–					
3. Trait anxiety	2.28 (0.57)	2.31 (0.63)	0.077	0.277*	–				
4. No room for expression	–0.01 (1.05)	0.01 (0.97)	0.011	0.469**	0.272*	–			
5. Criticism of ideas	0.08 (0.98)	–0.08 (1.03)	–0.293*	–0.051	–0.113	0.000	–		
6. Fear of evaluation	0.18 (1.06)	–0.18 (0.92)	–0.060	0.114	0.038	0.000	0.000	–	
7. Facilitator intervention	4.70 (1.61)	4.41 (1.62)	0.192	0.359**	–0.107	0.043	–0.132	–0.049	–

*p < *0.050*, **p < *0.010*.

productivity, t (52) = −1.20, p = 0.236, 95% CI [−2.77 0.70]. Further checks suggested that these findings could not be explained by facilitator intervention. That is, an independent-samples t-test showed no significant difference between brainstorming with a social robot facilitator (M = 4.70, SD = 1.61) and a human facilitator (M = 4.41, SD = 1.62), for the number of ideas the facilitator brought in, t (52) = 0.67, p = 0.504, 95% CI [−0.59 1.18]; and no significant correlation was found between facilitator intervention and productivity, r (54) = 0.192, p = 0.163. These findings suggest no evidence for a difference in productivity when people brainstorm with a social robot facilitator, compared to when they brainstorm with a human facilitator.

To explore whether brainstorming with a social robot facilitator, compared to brainstorming with a human facilitator, might increase productivity due to its effects on the relationship between state social anxiety and evaluation apprehension, and whether this is moderated by trait social anxiety, further analyses were conducted. That is, additional independent-samples t-tests were conducted with facilitator type as the independent variable, and social anxiety and the three evaluation apprehension factors as the dependent variables. Furthermore, correlations were calculated to explore whether the expected relationships between social anxiety, evaluation apprehension, and productivity could be confirmed. Combined, significant results could justify further exploration by means of (moderated) mediation analyses (Hayes 2017).

The results showed no significant difference between brainstorming with a social robot facilitator (M = 1.99, SD = 0.46) and a human facilitator (M = 1.86, SD = 0.44), for state social anxiety, t (52) = 1.08, p = 0.286, 95% CI [−0.11 0.37]. This finding was not likely to be unduly influenced by sampling errors. That, is an independent-samples t-test showed no significant difference between brainstorming with a social robot facilitator (M = 2.28, SD = 0.57) and a human facilitator (M = 2.31, SD = 0.63), for trait social anxiety, t (52) = −0.20, p = 0.846, 95% CI [−0.36 0.30]. As a consequence, trait anxiety does not moderate the relationship between facilitator type and state social anxiety. Further checks showed a significant positive correlation between facilitator intervention and state social anxiety, r (54) = 0.359, p = 0.008.

Regarding the three evaluation apprehension factors, the results showed no significant difference between brainstorming with a social robot facilitator (M = −0.01 SD = 1.05) and a human facilitator (M = 0.01, SD = 0.97), for no room for expression, t (52) = −0.11, p = 0.917, 95% CI [−0.55 0.49]; between brainstorming with a social robot facilitator (M = 0.08, SD = 0.98) and a human facilitator (M = −0.08, SD = 1.03), for criticism on ideas, t (52) = 0.56, p = 0.576, 95% CI [−0.40 0.70]; nor between brainstorming with a social robot facilitator (M = 0.18, SD = 1.06) and a human facilitator (M = −0.18, SD = 0.92), for fear of evaluation, t (52) = 1.37, p = 0.178, 95% CI [−0.18 0.87]. Note that the Shapiro-Wilk tests showed that the data of no room for expression and feature of evaluation were not normally distributed (p < 0.050). Therefore, emphasis must be placed on the bootstrapped 95% confidence intervals, rather than on the p-values.

The results did, however, show a significant positive correlation between state social anxiety and the evaluation apprehension factor no room for expression, r (54) = 0.469, p < 0.001; and a significant negative correlation between the evaluation apprehension factor criticism on ideas and productivity, r (54) = −0.293, p = 0.032.

Given these results further exploration by means of (moderated) mediation analyses is unlikely to provide further insight into the results. Therefore, these were not conducted (Hayes 2017). These findings suggest that, at least to some extent, the expected relationship between social anxiety, evaluation apprehension, and productivity was replicated. Brainstorming with a social robot facilitator, compared to a human facilitator, however, did not appear to influence this relationship in any way.

DISCUSSION

The presented study was conducted to take a first look at how brainstorming with a social robot facilitator compares to brainstorming with a human facilitator.

Summary and Interpretation of the Results

The results showed no evidence that brainstorming with a teleoperated social robot facilitator, compared to brainstorming with a human facilitator, influenced productivity. Where previous studies found positive effects of social robot facilitation compared to other technologies, such as an iPad application (Ali et al., 2021), PowerPoint presentation (Kahn et al., 2016) or a social robot that was turned off (Alves-Oliveira et al., 2020), the present study thus adds no evidence indicative that robot facilitation led people to generate more ideas than human facilitation. However, participants also did not generate fewer ideas with a robot than with a human facilitator. When a human facilitator is unavailable, or undesirable, a social robot might be a suitable replacement, provided that it can be programmed to facilitate brainstorming autonomously.

The results also showed no evidence that brainstorming with a social robot facilitator, compared to with a human facilitator, increased productivity due to its effects on the relationships between social anxiety and evaluation apprehension. This finding adds to previous work, which suggested that when people anticipate to collaborate on a task, people with a strong, compared to a weak disposition to experience social anxiety prefer to work with a social robot rather than a with human collaborator (Nomura et al., 2020). In the present study, participants actually worked with the social robot, but this had no notable effect on state social anxiety, nor was this effect moderated by trait social anxiety. Actually working with a social robot may thus not affect social anxiety, at least not within the context of an appropriately facilitated brainstorm, in this case by a professional facilitator *via* teleoperation. Further conjectures about subsequent effects on productivity *via* evaluation apprehension were therefore by extension also inaccurate.

The results did confirm, at least partly, the general theoretical assumptions about the relationships between social anxiety, evaluation apprehension, and productivity during brainstorming (**Table 2**). Social anxiety positively influenced participants' experience that there was no room for expression, and experienced criticism on their ideas negatively affected productivity (Leary 1983; Bordia et al., 2006). Moreover, a stronger disposition to experience social anxiety, led participants to experience more state social anxiety during the brainstorm (Spielberger 1966). Trait and state anxiety, however, did not influence productivity. Although this seems to contradict Camacho and Paulus's (1995) findings, their study was about brainstorming with peers rather than with a facilitator. Instead, the present study showed that the facilitator shared more ideas to keep the brainstorm going with participants that experienced more state social anxiety, compensating rightly for their reduced productivity (Sanders and Stappers 2008). The general psychological mechanism by which a social robot facilitator was thought to affect productivity, was therefore at least partially confirmed. It was just that no evidence was found of an effect of brainstorming with a robot facilitator, compared to a human facilitator, on these relationships between trait and state anxiety, evaluation apprehension, and productivity during brainstorming.

Limitations and Future Research

As with any first look, there are limitations that need to be taken into account when interpreting and building upon the results.

Firstly, next to any limitations introduced by the modest sample size, it may have been the case that the social anxiety experienced during the brainstorm was not sufficiently strong to lay bare effects of brainstorming with a social robot thereon. The scores on the trait and state anxiety questionnaires were low, indicating on average slight disagreement with statements indicative of social anxiety. This limits the generalizability of the results. Even so, if people with high trait and state social anxiety are the only demographic for which brainstorming with a social robot facilitator may be advantageous, this may not provide a strong case for investing in further research and development in social robot facilitators. Before such conclusions can be drawn, however, it may be advantageous to do further exploratory testing, a second look if you will, that includes further variables that may affect the relationship between human and robot facilitation, such as variation in level of training, facilitation styles, different group sizes, perceived robot autonomy, and online vs. offline differences.

Secondly, analysis of the evaluation apprehension scale revealed that three separate constructs were measured (**Table 1**). Although there were relationships between state social anxiety and no room for expression, and between criticism on ideas and productivity, these factors showed that state anxiety and productivity could not be correlated directly *via* the mechanisms that underlie evaluation apprehension. The imposed reliance on factor analysis, here, rather being able to rely on more in-depth theory to tease out the cause-and-effect relationships between social anxiety, evaluation apprehension, and productivity, threatens the study's internal and construct validity; which is further threatened by the resultant reliance on a non-simple factor structure (items 1 and 7), inclusion of factor loadings of close but inverted intensity (item 3), a factor expressed by a single item (factor "fear of evaluation" and item 6), and deviations from normality (factors "no room for expression" and "fear of evaluation") (e.g., Sellbom and Tellegen, 2019; Thurstone, 1947). See also **Table 1**. Combined with the fact that only one type of robot form was tested (the SoftBank Robotics NAO v5), further work could benefit from testing the effects of dedicated robotic forms and behaviors on the precise psychological mechanisms that drive the relationships between social anxiety, evaluation apprehension and productivity during brainstorming.

Thirdly, it must be noted that relying on the Wizard-of-Oz method threatens the study's external validity, because it remains to be seen whether the AI of a social robot can be developed to effectively deliver our brainstorm facilitation protocol. Although the Wizard-of-Oz method is widely used in research on the efficacy of brainstorming or doing other types of creative work with a social robot facilitator (Kahn et al., 2016; Alves-Oliveira et al., 2020), more research is needed to develop the computational backbone of social robot facilitators. In this regard, researchers such as Ali et al. (2021) are leading the way.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the TSHD Research Ethics and Data Management Committee, Tilburg University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed to developing the theory, method and article, and approved the submitted version. JG collected the data. AR conducted the statistical analysis.

REFERENCES

Abraham, Anna. (2018). *The Neuroscience of Creativity*. Cambridge University Press. doi:10.1017/9781316816981

Ali, S., Park, H. W., and Breazeal, C. (2021). A Social Robot's Influence on Children's Figural Creativity during Gameplay. *Int. J. Child-Computer Interaction* 28, 100234. doi:10.1016/j.ijcci.2020.100234

Alves-Oliveira, P., Arriaga, P., Cronin, M. A., and Paiva, A. (2020). "Creativity Encounters between Children and Robots," in Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, 379–388. doi:10.1145/3319502.3374817

Bartneck, C., and Forlizzi, J. (2004). A Design-Centred Framework for Social Human-Robot Interaction. " Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, 591–594. doi:10.1109/roman.2004.1374827

Bolin, A. U., and Neuman, G. A. (2006). Personality, Process, and Performance in Interactive Brainstorming Groups. *J. Bus Psychol.* 20 (4), 565–585. doi:10.1007/s10869-005-9000-7

Bordia, P., Irmer, B. E., and Abusah, D. (2006). Differences in Sharing Knowledge Interpersonally and via Databases: The Role of Evaluation Apprehension and Perceived Benefits. *Eur. J. Work Organizational Psychol.* 15 (3), 262–280. doi:10.1080/13594320500417784

Breazeal, C. (2011). Social Robots for Health Applications. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA. IEEE, 5368–5371. doi:10.1109/IEMBS.2011.6091328

Camacho, L. M., and Paulus, P. B. (1995). The Role of Social Anxiousness in Group Brainstorming. *J. Personal. Soc. Psychol.* 68 (6), 1071–1080. doi:10.1037/0022-3514.68.6.1071Mabel

Davis, N., Hsiao, C.-P., Popova, Y., and Magerko, B. (2015). An Enactive Model of Creativity for Computational Collaboration and Co-creation. In Springer Series on Cultural Computing. Springer, 109–133. doi:10.1007/978-1-4471-6681-8_7

Diehl, M., and Stroebe, W. (1987). Productivity Loss in Brainstorming Groups: Toward the Solution of a Riddle. *J. Personal. Soc. Psychol.* 53 (3), 497–509. doi:10.1037/0022-3514.53.3.497

Field, Andy. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Washington, DC: Sage Publications.

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A Survey of Socially Interactive Robots. *Robotics Autonomous Syst.* 42 (3–4), 143–166. doi:10.1016/S0921-8890(02)00372-X

Frich, J., MacDonald Vermeulen, L., Remy, C., Biskjaer, M. M., and Dalsgaard, P. (2019). "Mapping the Landscape of Creativity Support Tools in HCI," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–18. doi:10.1145/3290605.3300619

Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford publications.

Heimberg, R. G., Mueller, G. P. Craig. S. Holt, Debra. A. Hope., Holt, C. S., Hope, D. A., and Liebowitz, M. R. (1992). Assessment of Anxiety in Social Interaction and Being Observed by Others: The Social Interaction Anxiety Scale and the Social Phobia Scale. *Behav. Ther.* 23 (1), 53–73. doi:10.1016/S0005-7894(05)80308-9

Isaksen, Scott. G., Brain, K., and Treffinger, Donald. J. (2010). *Creative Approaches to Problem Solving: A Framework for Innovation and Change*. Thousand Oaks, CA: Sage Publications.

Kahn, P. H., Kanda, T., Ishiguro, H., Gill, B. T., Shen, S., Ruckert, J. H., et al. (2016). Human Creativity Can Be Facilitated through Interacting with a Social Robot. In ACM/IEEE International Conference on Human-Robot Interaction, 2016-April. 173–180. doi:10.1109/HRI.2016.7451749

Leary, M. R. (1983). Social Anxiousness: The Construct and its Measurement. *J. Personal. Assess.* 47 (1), 66–75. doi:10.1207/s15327752jpa4701_8

Nomura, T., Kanda, T., Suzuki, T., and Yamada, S. (2020). Do People with Social Anxiety Feel Anxious about Interacting with a Robot? *AI Soc.* 35 (2), 381–390. doi:10.1007/s00146-019-00889-9

Osborn, Alex. F. (1957). *Applied Imagination*. New York, NY: Scribner.

Paulus, P. B., and Yang., H.-C. (2000). Idea Generation in Groups: A Basis for Creativity in Organizations. *Organizational Behav. Hum. Decis. Process.* 82 (1), 76–87. doi:10.1006/obhd.2000.2888

Pinsonneault, A., Barki, H., Gallupe, R. B., and Hoppen, N. (1999). Electronic Brainstorming: The Illusion of Productivity. *Inf. Syst. Res.* 10 (2), 110–133. doi:10.1287/isre.10.2.110

Pot, E., Monceaux, J., Gelin, R., Maisonnier, B., and Robotics, Aldebaran. (2009). Choregraphe: A Graphical Tool for Humanoid Robot Programming. In Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, 46–51. doi:10.1109/ROMAN.2009.5326209

RIVM (2018). Factsheet Mentale Gezondheid Van Jongeren: Enkele Cijfers En Ervaringen. *Rijsinstituut Voor Volksgezondheid En Milieu*. Available at: https://www.rivm.nl/documenten/factsheet-mentale-gezondheid-van-jongeren-enkele-cijfers-en-ervaringen .

Sanders, E. B.-N., and Stappers, P. J. (2008). Co-Creation and the New Landscapes of Design. *CoDesign* 4 (1), 5–18. doi:10.1080/15710880701875068

Sawyer, R. Keith. (2011). *Explaining Creativity: The Science of Human Innovation*. Oxford University Press. doi:10.1093/oxfordhb/9780199204540.003.0003

Scassellati, B., Henny Admoni, Henny., and Matarić, M. (2012). Robots for Use in Autism Research. *Annu. Rev. Biomed. Eng.* 14, 275–294. doi:10.1146/annurev-bioeng-071811-150036

Sellbom, M., and Tellegen, A. (2019). Factor Analysis in Psychological Assessment Research: Common Pitfalls and Recommendations. *Psychol. Assess.* 31 (12), 1428.

Spielberger, C. D. (1966). Theory and Research on Anxiety. *Anxiety Behav.* 1 (3), 3–20. doi:10.1016/b978-1-4832-3131-0.50006-8

Thurstone, L. L. (1947). *Multiple-Factor Analysis; A Development And Expansion of the Vectors of Mind*. Chicago, IL: University of Chicago Press.

Warr, A., and O'Neill, E. (2005). "Understanding Design as a Social Creative Process," in Proceedings of the 5th Conference on Creativity & Cognition, 118–127. doi:10.1145/1056224.10562422*005*

Zawieska, K. (2014). Social Robots: Fostering Creativity through the Illusion of Life. In Workshop on Humanoid Robots and Creativity at the IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS2014).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 28 June 2021
doi: 10.3389/frobt.2021.662036

Embodiment in 18th Century Depictions of Human-Machine Co-Creativity

Anna Kantosalo[1], Michael Falk[2]* and Anna Jordanous[3]**

[1]Department of Computer Science, School of Science, Aalto University, Espoo, Finland, [2]School of English, University of Kent, Canterbury, United Kingdom, [3]School of Computing, Cornwallis South, University of Kent, Canterbury, United Kingdom

Artificial intelligence has a rich history in literature; fiction has shaped how we view artificial agents and their capacities in the real world. This paper looks at embodied examples of human-machine co-creation from the literature of the Long 18th Century (1,650–1,850), examining how older depictions of creative machines could inform and inspire modern day research. The works are analyzed from the perspective of design fiction with special focus on the embodiment of the systems and the creativity exhibited by them. We find that the chosen examples highlight the importance of recognizing the environment as a major factor in human-machine co-creative processes and that some of the works seem to precede current examples of artificial systems reaching into our everyday lives. The examples present embodied interaction in a positive, creativity-oriented way, but also highlight ethical risks of human-machine co-creativity. Modern day perceptions of artificial systems and creativity can be limited to some extent by the technologies available; fictitious examples from centuries past allow us to examine such limitations using a Design Fiction approach. We conclude by deriving four guidelines for future research from our fictional examples: 1) explore unlikely embodiments; 2) think of situations, not systems; 3) be aware of the disjunction between action and appearance; and 4) consider the system as a situated moral agent.

Keywords: human-machine co-creativity, embodiment, creativity, design fiction, literature, digital humanities, computational creativity

OPEN ACCESS

Edited by:
Vasanth Sarathy,
Tufts University, United States

Reviewed by:
Luisa Damiano,
Università IULM, Italy
Todd Lubart,
Université de Paris, France
Jeffrey Nickerson,
Stevens Institute of Technology, United States

****Correspondence:***
Anna Kantosalo
anna.kantosalo@aalto.fi
Michael Falk
M.G.Falk@kent.ac.uk
Anna Jordanous
A.K.Jordanous@kent.ac.uk

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *31 January 2021*
Accepted: *08 June 2021*
Published: *28 June 2021*

Citation:
Kantosalo A, Falk M and Jordanous A (2021) Embodiment in 18th Century Depictions of Human-Machine Co-Creativity. Front. Robot. AI 8:662036. doi: 10.3389/frobt.2021.662036

1 INTRODUCTION

Tools for assisting creativity are becoming more commonplace. New systems utilizing artificial intelligence (AI) methods to empower the tools themselves to be creative are stepping in different fields, including robots for playing music (Hoffman and Weinberg, 2010; Weinberg et al., 2020) and singing (Miranda, 2008), sketching (Lin et al., 2020) and even fostering creativity in children (Ali et al., 2019). These co-creative robots represent technological progress in machine engineering, artificial intelligence as well as human-machine interaction.

The idea of machines assisting creativity precedes the current practical advancements by hundreds of years. Simple tools such as musical dice were invented in the 18th Century to enable musically ignorant persons to take part in writing music and received huge popularity in the contemporary intellectual climate fuelled by rationalism (Hedges, 1978). Provided with a pre-composed set of musical fragments, typically musical phrases that fit certain melodic or harmonic constraints, people could construct coherent musical compositions by using dice rolls to select fragments from the set. These musical dice games have acted as inspirations for more modern AI based approaches to interactively compose music with computers (see e.g., Lin et al., 2015).

In this paper we explore examples of co-creative systems from Eighteenth-Century literature in different areas of artistic creativity. We have focused on robotic and other physical systems to emphasize the embodied aspects of human-machine interaction. Like the example of musical dice games we expect these systems to fuel the imagination of modern-day researchers. In the same way that present-day design fictions combine science fiction and design research (Sterling, 2005; Bleecker, 2009; Dunne and Raby, 2013; Blythe, 2014), these older fictions depict potential solutions for solving complex creativity related problems and offer us a new perspective for questioning our design solutions and research approaches. Bruce Sterling, inventor of the term "design fiction", argues that design fictions aim to "suspend disbelief about change" (quoted in (Blythe, 2014, p. 3)). By taking old speculations about human-machine co-creation seriously, we may discover that new kinds of human-machine co-creation are possible.

2 MATERIALS AND METHODS

This work aims to develop our understanding of embodied creativity, by increasing our knowledge of how human-machine co-creativity has been understood in the past. We analyze five examples, selected by our literary expert author as representative of human-machine co-creativity in the Long 18th Century (c. 1,650–1850): two Romantic poems about aeolian harps (harps played by the wind): Samuel Taylor Coleridge's "The Eolian Harp" (1795, rev. version 1817, in Coleridge and Keach (1997)) and Eduard Mörike's "An eine Äolsharfe" (1837, in Mörike (1838)); E. T. A. Hoffman's "Automata" (1814), a tale containing a multitude of automatic musical systems and a humanoid question/answer machine; the creativity thinking aids featured in Laurence Sterne's novel *Tristram Shandy* ([1759–67] 2009); the self-conscious Hackney Coach from Dorothy Kilner's *Adventures of a Hackney Coach* (1781); and the artificial man Homunculus in Goethe's play *Faust: Der Tragödie Zweite Teil* (1832). Like contemporary design fictions, these literary design fictions take a variety of forms (poetry, fiction, drama).

We have chosen to focus on the Long 18th Century (c. 1650s–1850s) because it marked a watershed in the history of AI (Riskin, 2016). At the beginning of this period, René Descartes set the question of AI on a new footing with his mind-body dualism; by the end of this period, Mary Shelley's *Frankenstein* (1818) had spawned a powerful myth that still dominates the way AI is imagined today. In between Descartes and Shelley there was a period of great imaginative freedom, when authors experimented with many different kinds of fictional AIs.

Descartes had shattered the Thomistic consensus that mind and matter were interfused, and that only God could create new forms of life. He argued that the body was a mere machine, and that functions of life such as ingestion and sense-perception could be explained by the mechanical workings of matter; the soul was utterly separate from the body, and was responsible for abstract thought alone (Descartes, 1988, 64–65). Later in the period, radical materialists such as Julian Offray de La Mettrie (1996) would dispense with the soul, arguing that thought was also just a function of the body's machine. These arguments made it possible to believe that scientists might create a living or intelligent machine using the laws of physics alone, and fired the imaginations of writers and inventors alike. In our chosen examples, all sorts of objects are imagined as potentially intelligent: from bowling-greens and Pentagraphs to harps and hackney-coaches. These visions of AI can seem strange and eccentric to the twenty-first-century reader. This is precisely why they are worth considering.

By the end of the Long Eighteenth-Century, speculations about AI had become commonplace, and the marvellous automata that had dazzled the European public had begun to lose their allure for an intellectual or scientific audience (Hankins and Silverman, 1999, 213–216). With her blockbuster novel *Frankenstein*, Mary Shelley simultaneously revived public interest in AI and sent the discussion in a new direction. The myth of the rebellious superintelligence was born. In the 19th Century, novelists such as Samuel Butler (1872) and George Eliot (1879) extended Shelley's ideas about how AI might evolve beyond its human creators. In our own time, AI theorists such as Kurzweil (2006), Bostrom (2014), Tegmark (2018) and (Russell et al., 2019) have attempted to bring the *Frankenstein*-myth into the scientific mainstream (Falk, 2021). Meanwhile rebellious super-intelligent AIs remain a staple of contemporary science fiction. By looking back to the Long 18th Century, before the *Frankenstein*-myth set in, we hope break open the scientific imagination, and open up new ways of thinking about the roles AI might play in human life.

In our search for new ways of thinking about the roles of AI, we choose to focus on human-machine co-creativity. In this paper human-machine co-creativity is considered as a collaborative activity between a human and a machine driven toward an artistic goal. In human–computer co-creativity, co-creation is often understood as the creation of artifacts via the interaction of different initiatives (Yannakakis et al., 2014), sharing of creative responsibility (Kantosalo et al., 2014) or blending the machine into the human creative process (Davis, 2013). The term encompasses various different ways of organizing the co-creative process and the human and the machine can play different kinds of roles (for example, see the classifications of such roles by Kantosalo and Jordanous (2020); Lubart (2005)) or contribute to the creative process in different ways (Kantosalo and Takala, 2020). The style of human–computer co-creative interfaces is often similar to mixed-initiative interaction (Allen et al., 1999; Horvitz, 1999) for this reason, human–computer co-creativity is sometimes referred to as mixed-initiative co-creativity (Yannakakis et al., 2014) and the related interfaces as mixed-initiative creative interfaces (Deterding et al., 2017). In general human–computer co-creative systems can express various degrees of co-creative intent on a spectrum from full human intent to full computational intent (Deterding et al., 2017).

TABLE 1 | Embodied interactions in each example text.

Work	Artificial agent	Agent's embodiment	Human's embodiment
Coleridge's "Eolian Harp" (1795)	An aeolian harp	String instrument played by the wind	The human poet (Coleridge) interacts using sense of hearing
Mörike's "An eine Äolsharfe" (1837)	An aeolian harp	String instrument played by the wind	The human poet (Mörike) interacts using sense of hearing
Goethe's *Faust II* (1832)	Homunculus	Tiny artificial human enclosed in a glass phial; glows and can read minds	Human characters converse with Homunculus or have their minds read
Hoffman's "Automata" (1814)	The Talking Turk	Clockwork fortune-teller with power of speech and mysterious inner mechanism	Human characters ask the Turk questions, hear its answers and peer into its mechanism
	Artificial performers	When activated by a human, the performers create music	Humans activate the artificial performers, play alongside them, and listen as the audience
Kilner's *Hackney Coach* (1781)	A hackney coach	Coach can see and hear its immediate surroundings, but does not control its own movements	Humans drive or ride on the coach, unwittingly providing the material for its narrative
Sterne's *Tristram Shandy* (1759–67)	Mechanical writing-aides	Audio or optical devices that change the appearance of an observed object	Human narrator imagines using different aides to gain different insights into his human characters
	The bowling green	A bowling green of soft earth that is shaped into a scale model of key battlefields in the nine years' war	Human characters manually update the bowling green model as news arrives from the front

The selected fictitious examples are analyzed as design fictions for variations in levels of embodiment and creativity. Embodiment was selected as a perspective, since it builds a bridge between contemporary artificial intelligence research and eighteenth-century philosophy. Eighteenth-century scientists such as Jacques de Vaucanson and Wolfgang von Kempelen argued that the achievement of AI would require simulation of bodily systems (Riskin, 2003; Riskin, 2016). These arguments appeared to be justified at the time by extraordinary advances in the design of mechanical calculators, in the construction of "automata" (clockwork robots that replicated bodily movements), and in fields such as optics and acoustics. Meanwhile philosophers such as Étienne Bonnot de Condillac (1984) argued that cognition of external objects would be possible only for an embodied being with a sense of touch. These eighteenth-century arguments foreshadow debates in phenomenology, cognitive science and neuromorphic computing today. It was a period of considerable speculation about the possibility of artificial intelligence, and the literature of the period may contain useful lessons for scientists today.

2.1 Selected Works for Analysis

As shown in **Table 1**, the examples chosen for this work represent a variety of co-creative systems ranging from the very tool-like systems presented in *Tristram Shandy* (Sterne, 2009), through to more genuinely co-creative examples such as the mechanical musicians in E. T. A. Hoffman's "Automata" (1814) or the almost autonomous Hackney Coach in (Kilner, 1781). The examples also represent different kinds of embodiment, from the utterly non-human aeolian harps in Coleridge and Mörike's poems (Coleridge and Keach, 1997; Mörike, 1838), to Hoffman's humanoid Talking Turk (Hoffman, 1957), and the essentially disembodied Homunculus in Goethe's *Faust* (Goethe, 1832).

The first two examples are the aeolian harp poems of Samuel Taylor Coleridge and Eduard Mörike. Aeolian harps were popular stringed instruments of the 18th and 19th centuries. They could take various forms, but all aeolian harps were harps designed to be played by the wind rather than human fingers. They were of particular interest to early researchers in acoustics, who were perplexed by the harps' peculiar creative properties: when a harp's string is plucked by a human, it can only produce one note, but when played by the wind, it can produce a great variety of different notes (Hankins and Silverman, 1999, ch. 5). Poets like Coleridge and Mörike were also interested in the harps' creative properties, but interpreted the harp in a more mystical way. Coleridge, for instance, seems to have believed that the harp, the wind, and its human listener all participated in a shared consciousness:

> And what if all of animated nature
> Be but organic Harps diversely fram'd,
> That tremble into thought, as o'er them sweeps,
> Plastic and vast, one intellectual Breeze,
> At once the Soul of each, and God of all?
> (Coleridge and Keach, 1997)

In Mörike's poem, there is a similar ambiguity. As the poet listens to the harp, it is hard to distinguish whether the emotions of the music are the harp's emotions, the poet's emotions, or emotions that are latent in the situation as a whole. We see these as poetic examples of extended consciousness.

In Part Two of Johann Wolfgang von Goethe's *Faust*, we encounter a more typical fictional AI: the creature Homunculus. Homunculus is an artificial man created by Wagner, a scientist and alchemist. What makes Homunculus unusual, especially when compared to more famous fictional AIs such as Frankenstein's monster, is his embodiment. His human body is minuscule, and contained within a fragile glass phial. He doesn't seem to make any use of his human body parts. Instead of walking, he floats in mid-air. He also has the ability to glow, read human thoughts, and later, absorb himself into other beings. In some regards, Homunculus resembles a disembodied software program more than an embodied human, despite his human appearance, and looks forward to the disembodied AIs of cyberpunk classics such as Masamune (2009); Gibson (2016).

The artificial agents in E.T.A. Hoffman's story, "Automata", are more down-to-earth, because they are based on actual automata that were built and exhibited in Eighteenth-Century Europe. At the beginning of the story, the main characters encounter the Talking Turk, a clockwork question-answering system of stereotypically Turkish appearance. The Turk answers users' questions, and is cleverly designed to defy users' attempts to work out how it operates. The other automata in the story are a group of clockwork musicians, based on some famous examples by the French artificer Jacques de Vaucanson (Riskin, 2016, ch. 6). The main characters meet the creator of these musicians. The creator turns them on, and plays a piece with them on his piano. One useful feature of this example is that the main characters disagree about whether any of the automata are truly intelligent or creative.

Dorothy Kilner's *Adventures of a Hackney Coach* (1781) is an "it novel" or "novel of circulation". This was a popular genre of fiction in the late 18th Century, in which an inanimate character such as a coin, atom or statue would narrate their adventures in the world (Bellamy, 2007). In Kilner's novel, the narrator is a Hackney Coach, who is driven around London and the surrounding area picking up passengers from different social classes. The novel is told from the Hackney Coach's perspective, as it recounts the conversations of its passengers and describes the different places and events it visits. There are two key creative collaborations: the Coach collaborates with its passengers, who unwittingly provide the material for the story, and the Coach collaborates with the human reader, to whom the narrative is addressed.

Sterne's novel *Tristram Shandy* (Sterne 2009) is the fictional autobiography of the main character, Tristram Shandy. It is a famously experimental and unusual novel. Although it is ostensibly the story of Tristram's life, he gets so sidetracked describing the lives of his father, Walter Shandy, and his Uncle Toby, that never manages to narrate more than the first few years of his own life. The book also contains many digressions, where Tristram discusses the art of novel-writing and other mostly irrelevant topics.

Tristram Shandy includes two interesting examples of co-creative machines. In Volume 1, Chapter 23 the narrator describes several imaginary writing aids that allow the writer to develop character depictions. Momus's glass is a device installed on a character's chest, which enables the writer to perceive the inner workings of the character's soul as if through a window. Other writing aids include musical instruments that play the characters' emotions, a Pentagraph, a mechanical device that exactly replicates the movements of the human writer's pen, and the "Hobby-Horse", which Tristram uses to reveal a character's driving obsession. For clarity, we have focused on Momus' glass, the most extreme and also best-described of Tristram's imaginary co-creative machines. The second main co-creative system is the bowling green used by Uncle Toby, which Tristram describes in particular in Volume 2, Chapter 1, and Volume 6, Chapters 21–23. The bowling green is a massive model of the Nine Years' War, and, ironically enough, it is Uncle Toby's "Hobby Horse", linking it to the first set of co-creative machines. We look at this model as a physical co-creative environment in which Uncle Toby acts and analyses various events of the war.

2.2 Design Fiction

How can two hundred year old fictional texts inform scientific research today? None of the examples we have chosen describe a scientific process by which a creative artificial agent could be made. Indeed, some of the examples are impossible. Dorothy Kilner's intelligent Hackney Coach, for example, is able to see and hear events in its immediate surroundings even though it apparently lacks any sensory organs, and its personality is notably human and English despite the fact it is a coach. Rather than explaining how this kind of intelligent agent exists, the novel takes the agent for granted and explores its implications.

We therefore propose to interpret these texts as *design fictions* (Sterling (2005); Bleecker (2009); Dunne and Raby (2013); Blythe (2014)). The purpose of design fictions is not to show scientists how to solve a problem, but rather to help scientists determine what problem they should try to solve. Design fictions achieve this purpose by simulating the experience of interacting with new technologies. Instead of describing how a particular technology functions, or describing the process of manufacture, design fictions presuppose that such a technology exists, and portray what interacting with it would be like. A design fiction is typically in the form of a film, novel, poem, play or art installation. When we read or view such a fiction, we imagine ourselves in a new world, where a new technology exists, and can feel what it might be like to live with such a device. We imagine the future 'from the inside', and gain an intuitive, embodied, subjective understanding of which technologies we might desire to have or wish to avoid (Burdick, 2019). In this way, design fiction offers two key benefits to scientists: goalposts and warning signs. Goalposts: by painting a vivid picture of humanity's possible futures with technology, design fiction can fire scientists' imaginations and widen the search space. Warning signs: by allowing us to explore the "implications" rather than simply the "applications" of new technologies, design fiction can help scientists determine the moral and ethical implications of their research (Dunne and Raby, 2013, p. 49).

These historical texts were not necessarily intended as "design fictions", but by focusing on the interaction between the human and non-human agents, we can read them as if they were. To read these texts as design fictions, we focus specifically on embodied interaction. How does the embodiment of the fictional artificial agent affect its cognition of the world and its interaction with human beings? And likewise, how does the embodiment of human agents affect their interaction with the artificial agents? In Mörike and Coleridge's poems, for example, the artificial agent is an aeolian harp. Such harps are large stringed instruments that are played by the wind. Due to this embodiment, they must be placed outside or in a window, where the wind blows, and they produce output in the form of sound. The agent is therefore stationary, under the wind's influence, and surrounded by outdoor scenes. To interact with the harp, the human agent must go outside, wait for the wind, and listen with their ears. Thus the embodiment of the harp and the human intersect to produce a particular kind of interaction. The poems describe this interaction in vivid language, recreating the experience of listening to the aeolian harp and offering the reader an intuitive understanding of what value such an interaction could have.

2.3 Embodiment and Creativity

Viewpoints on embodied creativity and collaboration, and on embodied artificial intelligence more generally, range across a number of notions (Chrisley, 2003; Ziemke, 2015). Competing theories about the role of embodiment in cognition range from theories of minimal embodiment, which look at the embodiment of humans through a reduced body without a brain, to views of embodiment that connect the body with the environment, or allow it to absorb tools to extend its embodiment, and all the way to radical views of embodiment viewing perception as an action oriented concept shaping most cognitive processes (Gallagher, 2011), or behavior-based robotics (Brooks, 1991). Dreyfus (1979) argued that some form of embodiment was necessary for intelligence, echoing the arguments of eighteenth-century materialists such as Condillac and La Mettrie.

Very few works of human–computer co-creativity address aspects of embodiment specifically or take a stance on different theories of embodiment. Embodiment is briefly discussed in some works featuring robotic collaborators, such as an investigation of task-transfer (Fitzgerald et al., 2017), and the works of Saunders, Gemainboeck and their colleagues (e.g., Saunders et al., 2010, 2013; Gemeinboeck and Saunders, 2013), which feature embodied robots that allow for shifting the balance of co-creative initiative toward machine initiative. A few theoretical works, discuss the extended mind theory (Bown, 2015), an enactivist theory for co-creation (Davis et al., 2015) and the wide acceptance of creativity as a situated activity within the field of computational creativity (Guckelsberger et al., 2017), but there appears to be no unified view of how to look at embodiment from the perspective of human–computer co-creativity. Therefore for this paper we wanted to find a disambiguation of embodiment that would more directly address creativity and collaboration with machines.

Dag Svanæs (2013) created a bridge between creativity and embodiment in his work investigating the role of embodiment in interactive technologies. Svanæs applies Merleau-Ponty's ideas about the lived body and embodied perception into analyzing interaction with technology. As a result he created three concepts "the feel dimension", "interaction gestalts" and "kinaesthetic thinking" which he used to discuss different kinds of digital products and interfaces. From the last concept he developed the idea of "kinaesthetic creativity", which discusses how a designer, embedded in a rich context through their lived body can use that experience to create new solutions to problems perceived in that moment.

This paper examines the effects the embodiment of the example systems may have on their creativity. In his paper, Svanæs (2013) shows how to use Merleau-Ponty's ideas selectively to perform formative analysis of interaction with a few examples ranging from abstract creative art to e-readers. Svanæs' examples focus on software oriented artifacts with tangible physical interfaces. To be able to compare fictitious robotic examples in a summative, systematic manner, we took the twelve key components Svanæs' derives from Merleau-Ponty's (1962) to support his user interaction concepts and turned them into comparison criteria. The twelve criteria are *active perception; perception shaped by the phenomenal field; directed perception; mediating perception through artifacts; whole body perception; the lived body; incorporating artifacts into the body; body schema; bodily space; skills acquisition; the dynamic nature of the body, tools and objects; and concrete and abstract movement*.

Following Svanæs' (2013) descriptions, the first concept, *active perception*, focuses on human perception as active uses of senses instead of passive reception of stimuli. The second concept, *perception shaped by the phenomenal field*, looks at how the individual background, such as experiences and training, affect human perception. The third concept, *directed perception*, looks at the intentions of the individual affecting what and how they perceive. The fourth concept, *mediating perception through artifacts*, looks how the body can adapt and extend its perceptional capacity through the use of artifacts, such as by a visually impaired person navigating with a stick. The fifth concept, *whole body perception*, looks at how the whole body can be used automatically to extend perceptional capacity, such as turning an object while visually perceiving it to take in various angles. The sixth concept, *the lived body*, considers the body as a general medium of presence in the world. The seventh concept, *incorporating artifacts into the body*, looks at assuming artifacts as part of the lived body, such as a person using a wheelchair. The eight concept, *body schema*, describes the "nonconscious knowledge" an individual has of their lived body and its potential actions in the world. The ninth concept, *bodily space*, considers the degrees of freedom the lived body has in the space. The 10th concept, *skills acquisition*, considers how an individual is able to "internalize external devices through learning". The 11th concept, *the dynamic nature of the body, tools, and objects*, looks at the changing contextual meaning and purpose of the body, tools and objects. The 12th concept, *concrete and abstract movement*, looks at movements "made naturally as part of a situation" and movements made for the purpose of movement.

We combine this analysis of embodiment with a separate analysis of creativity. This allows us to explicitly consider how embodiment and creativity interact in fictions of this period. In modern creativity research, creativity is characterized by many aspects (Jordanous and Keller, 2016), which vary in importance across domains. Over the years some authors such as Kantosalo and Takala (2020) have attempted to establish frameworks for describing human–computer co-creativity that take into account various theories of human creativity, including for example Glăveanu (2013) and Csikszentmihalyi's (1988) views of creativity as a socio–cultural act and Glăveanu's views of material affordances of the creative environment. But to our knowledge, there is no single definition for creativity that is adopted over others in human–computer co-creativity research. Therefore for our analysis we have attempted to

TABLE 2 | Components of Creativity, with definitions adapted from Jordanous and Keller (2016).

Component	Definition (adapted from Jordanous and Keller (2016))
Active involvement and persistence	Being actively involved; reacting to, deliberate Tenacity to persist, even at problematic points
Generation of results	Working toward some end target, or goal, or result Producing something that previously did not exist
Dealing with uncertainty	Coping with incomplete, missing or ambiguous information Element of risk/chance, lack of routine/pre-existing methods
Domain competence	Domain-specific intelligence, knowledge, expertise Recognizing problems and generating new ideas in that domain
General intellect	General intelligence and intellectual ability Flexible and adaptable mental capacity
Independence and freedom	Working independently with autonomy over actions/decisions Freedom to work, perhaps challenging cultural/domain norms
Intention and emotional involvement	Personal and emotional investment, immersion Intention/desire to perform a task, for fulfilment/enjoyment
Originality	A new product, or doing something in a new way Results that are unpredictable, unexpected, surprising
Progression and development	Movement, advancement, evolution during a process Some developmental progression in a domain/task
Social interaction and communication	Communicating and promoting work to others Mutual influence, feedback, collaboration
Spontaneity/Subconscious processing	Thoughts may inform a process subconsciously Reacting quickly and spontaneously when appropriate
Thinking and evaluation	Consciously evaluating several options Proactively selecting a decided choice from possible options
Value	Making a useful contribution valued by others End product is relevant and appropriate
Variety, divergence and experimentation	Generating different ideas to compare and choose from Multi-tasking during a process

find some more general definitions of creativity that would fit the variety of the literary samples examined here.

Creativity research commonly adopts a minimal "bifold" definition of creativity (Runco and Jaeger, 2012), but such a minimal definition makes it difficult to classify the myriad forms of creativity encountered in literary texts. Creativity is an example of an *essentially contested concept* (Gallie, 1955; Jordanous, 2012), in that by its nature, creativity resists full, complete and fixed definition (Corazza, 2016). The nature of creativity has been much discussed from multiple perspectives and various disciplines (Jordanous and Keller, 2016) (e.g., Guilford (1950); Gero (1996); Gabora (2005); Hennessey and Amabile (2010); Weisberg (2015), as a selection of a few different perspectives in a vast and multi-disciplinary area of research).

It is widely acknowledged that practical concerns drive us to adopt working definitions where necessary (Runco and Jaeger, 2012), definitions which others may see as partial or incomplete for their purposes. This paper provides a good example: we argue that the "standard definition" of creativity proposed by Runco and Jaeger (2012) is insufficient for this work, whereas many creativity scholars would find this definition adequate for their purposes. Instead we include in our considerations 14 components of creativity, which were derived by Jordanous and Keller (2016) from computational analysis of a corpus of seminal articles spanning a period of 60 years of research on creativity, from multiple disciplinary perspectives. We do not claim that these components form a conclusive and complete definition of creativity; however for practical purposes these components enable a more divergent, detailed and multi-faceted analysis of the literary and embodied context of this work, considering aspects such as generative ability and originality, as well as social interaction and communication. The components are listed and briefly defined in **Table 2**.

3 RESULTS

We conducted our analysis of the works such that relevant passages of the works would be read by two researchers separately after which the researchers discussed the different elements of embodiment and the different aspects of creativity in the examples.[1] Based on these discussions we compiled two tables which allowed for comparing and contrasting the different examples through these elements and also to examine whether the embodiment of the systems had any interesting connections to the creative capacities exhibited by the examples (**Tables 3**, **4**). In each case, the question was whether the relevant fictional agent was capable of the given component of embodiment, or the given component of creativity. For example, in none of the fictional examples did the artificial agent acquire a new skill, so all received a null score for "skills acquisition" (**Table 3**). In some cases, the situation was ambiguous. In Hoffmann's "Automata", for example, the Talking Turk seems to display "general intellect"

[1]The exception was *Faust*, which was read by the literary expert alone, and then discussed at length with the team.

TABLE 3 | Analysis of works by embodiment criteria adapted from Svanæs (2013).

	Mörike, "Äolsharfe"	Coleridge "Eolian Harp"	Kilner, "Hackney Coach"	Sterne, "Tristram Shandy"; bowling green	Sterne, "Tristram Shandy"; writing aides	Goethe, "Faust: Zweiter Teil"	Hoffmann, "Automata"; Talking Turk	Hoffmann, "Automata"; instruments
Active perception	x	x	x	—	—	x	x	—
Perception shaped by the phenomenal field	—	—	x	—	—	x	—	—
Directed perception	x	x	x	—	x	x	??	—
Mediating perception through artifacts	—	—	x	—	—	—	—	—
Whole body perception	x	x	—	—	—	x	—	—
The lived body	—	—	x	—	—	x	—	—
Incorporating artifacts into the body	x	x	??	x	x	x	—	x
Body schema	x	x	x	—	—	—	—	—
Bodily space	x	x	x	x	x	—	x	x
Skills acquisition	—	—	—	—	—	—	—	—
The dynamic nature of the body, tools and objects	—	—	—	x	x	—	—	—
Concrete and abstract movement	—	—	??	—	—	??	—	—

TABLE 4 | Analysis of works by creativity component.

	Mörike, "Aolsharfe"	Coleridge "Eolian Harp"	Kilner, "Hackney Coach"	Sterne, "Tristram Shandy"; bowling green	Sterne, "Tristram Shandy"; writing aides	Goethe, "Faust: Zweiter Teil"	Hoffmann, "Automata"; Talking Turk	Hoffmann, "Automata"; instruments
Active involvement and persistence	x	x	x	x	—	x	x	—
Dealing with uncertainty	—	—	x	—	—	—	—	—
Domain competence	—	x	x	—	—	—	x	—
General intellect	—	—	x	—	—	—	??	—
Generating results	x	x	x	x	—	x	x	—
Independence and freedom	x	x	??	—	—	x	x	—
Intention and emotional involvement	x	x	x	—	—	x	x	—
Originality	—	—	x	—	x	x	x	x
Progression and development	—	—	—	—	—	x	—	—
Social interaction and communication	x	x	x	x	—	x	x	—
Spontaneity and subconscious processing	x	x	x	—	—	x	x	—
Thinking and evaluation	??	??	x	—	—	x	??	—
Value	x	x	x	x	x	x	x	??
Variety, divergence and experimentation	x	x	??	??	x	x	x	—

because it is able to listen to and answer any question posed by a member of the public (**Table 4**). But the story is deliberately ambiguous as to whether the android is actually intelligent or is just a hoax: when the Turk talks and gestures, "die Rückwirkung eines denkenden Wesens unerläßlich schien" [the agency of some intelligent being seemed essential] (Hoffmann, 1957, vol. 6, p. 82). When we felt that an example was insuperably ambiguous in this way, we have placed a "??" in the relevant cell of the table.

4 DISCUSSION

Advocates of speculative design argue that design fictions "can play a significant role in broadening our conception of what is possible" (Dunne and Raby, 2013, p.162). Our examples could play this role, by helping scientists rethink core concepts of computational creativity, including embodiment, agency and creativity itself. These texts break down the usual template of creative activity: the human being. They suggest that radically non-human actants such as musical instruments, vehicles or even the ground may exercise certain kinds of creative agency. If scientists engage with these texts, they may rethink their assumptions about the form a creative system might take, invite new analysis of ethical and social implications, and open new lines of inquiry.

4.1 Concrete and Abstract Body: Varying Levels of Agency

Most of our examples describe creativity as an automatic or mechanical process, which does not require intellect or self-consciousness. In **Table 3**, for example, only a few of the examples display "skills acquisition", a "dynamic relationship between body, tools and objects", or the distinction between "concrete and abstract movement". This last component is particularly interesting in the cases of the Hackney Coach and Homunculus. Both of these artificial agents have human-level general intelligence, and are able to interact socially with human beings. The Hackney Coach composes a novel describing its life, while Homunculus converses with other characters. But neither of them seem to distinguish between unconscious "concrete" movement (e.g., the automatic movement of the fingers while typing) and conscious, intentional 'abstract' movement (e.g., carefully positioning the fingers to pick up a sharp object). The Hackney Coach moves passively, according to the pulling force of its horse and the direction of its driver. It is therefore capable *only* of "concrete", unconscious movement, as its wheels turn to accommodate the direction of the horse and driver. Homunculus, by contrast, moves intentionally, but his form of movement (levitation) requires no bodily action. In his case, he is capable of purely "abstract", intentional movement, but not of "concrete", unconscious movement. In this way, both these agents break down the distinction between "conscious" and "unconscious", at least as far as bodily movement is concerned.

This lack of bodily self-consciousness correlates with several components of creativity. Few of the agents "deal with uncertainty", possess "domain competence", have "general intellect", or "progress and develop" (**Table 4**). Their creativity is generally spontaneous and adventitious, rather than self-conscious and deliberate. In Coleridge and Mörike's poems, for instance, the aeolian harps create original music, decode emotional information that is encoded on the wind, and communicate it to human listeners who participate by providing the missing ingredient of self-consciousness. Mörike, for instance, describes how the wind:

> . . . säuselt her in die Saiten,
> Angezogen von wohllautender Wehmuth,
> Wachsend im Zug meiner Sehnsucht,
> Und hinsterbend wieder. (Mörike, 1838, p.52)
> [. . . rustles hither in the strings,
> Drawn by eloquent melancholy,
> Growing in the pull of my desire,
> And dying away again.]

The wind appears to feel some emotion, being itself "drawn" to the poet's melancholy. Meanwhile the poet responds emotionally to the music of the wind in the strings of the harp. The harp creates such music, and communicates such emotion, without apparently having any kind of intellect or consciousness.

This example raises the difficult problem of agency: How can we ascribe creative intentions or actions to the harp or the wind? In his own aeolian harp poem, Coleridge speculates that there is "one life, within us and abroad, — Which meets all motion and becomes its soul" (Coleridge and Keach, 1997, p. 87). If there is indeed such a global "soul", "life" or consciousness that pervades all things, then it would be perfectly possible for the wind to intend or to act. But such a "world-soul" is scarcely consistent with modern science. According to Riskin (2016, pp. 1–2), in contemporary physics and chemistry it is considered unacceptable to describe any physical system by ascribing agency to its components. Probabilities and causes are acceptable explanations, not decisions. This hesitancy over agency is quite palpable in the field of computational creativity. In a classic definition of the field, Wiggins (2006, p. 2) says that Computational Creativity is:

> The study and support, through computational means and methods, of behaviour exhibited by natural and artificial systems, which would be deemed creative if exhibited by humans.

This strange phrase, "would be deemed creative', indicates an insecurity at the heart of the field. If no agency can be ascribed to "natural and artificial systems", then how can a computer actually *be* creative? At best it can merely simulate or model creative activity.

This problem of agency, creativity and machinery is addressed explicitly in Hoffmann's "Automata", when the characters debate whether the mechanical musicians create genuine music. After hearing the mechanical musicians, the characters Ludwig and Ferdinand come to differing conclusions. Ludwig finds the automatons' music "zuwider" [repugnant], arguing that the agency of a human musician is required to create true music (Hoffmann, 1957, vol. 6, p. 105).[2] Ferdinand finds the artificial music beautiful, though agrees that it is inferior to human music. Interestingly, although Ludwig loathes the machines' music, he nonetheless claims that music has an ultimately nonhuman source:

> "Kann denn", ewiderte Ludwig, "die Musik, die in unserm Innern wohnt, eine andere sein als die, welche in der Natur wie ein tiefes, nur dem höhern Sinn erforschliches Geheimniss verborgen, und die durch das Organ der Instrumente nur wie im Zwange eine mächtigen Zaubers, dessen wir Herr worden, ertönt?" (Hoffmann, 1957, vol. 6, p. 107) ["Can it be," replied Ludwig, "that the music that lives within us is different to that which lies as a deep mystery in Nature, discoverable only by the highest sense, and which is expressed by instruments only under the compulsion of a mighty spell of which we are the masters?"]

For this reason, although the automaton musicians are not to his taste, Ludwig is more open to the music of Aeolian harps, and claims that a "höhere musikalische Mechanik" ["higher

[2]This aligns with modern-day interpretations of what it means to be creative when generating music (Jordanous and Keller, 2012): social communication and interaction, domain competence and intention/emotional involvement were found to be crucial factors, and Ludwig is questioning the ability of the automatons to engage with their music intentionally and with emotion.

mechanics of music"] is possible (Hoffmann, 1957, vol. 6, p. 105). Thus even this more sceptical, scientific text leaves open the possibility that nature itself may have creative properties that could be harnessed by a mechanical system.

This issue of agency is particularly prominent in *Tristram Shandy*, whose artificial agents lack not only self-consciousness, but also "independence and freedom", "intention and emotional involvement" and also, in one case "originality" (**Table 4**). In terms of embodiment, the agents also lack "active perception" and the "shaping force of the phenomenal field" (**Table 3**). The bowling green is entirely shaped by human hands, and has no sense organs of any kind, while the writing aids, such as Momus' glass, are essentially optic, acoustic or haptic instruments that provide the writer with novel input about the character they are trying to describe. Momus' glass provides a vision of the character's heart, but there is no suggestion that the glass actually *sees* the character's heart itself.

Nonetheless, in the novel both the writing aids and the bowling green frequently intervene in the plot and change the course of events, displaying a form of material agency which does not require any kind of awareness. In a mundane way, the bowling green's soil composition affects how well Uncle Toby's model operates:

> Nature threw half a spade full of her kindliest compost upon it, with just so much clay in it, as to retain the forms of angles and indentings—and so little of it too, as not to cling to the spade, and render works of so much glory, nasty in foul weather (Sterne, 2009, p. 342, p. 342).

This is the most obvious way the bowling green's embodiment affects its collaboration with humans, but throughout the novel, the bowling green acts in other, more surprising ways. It is directly implicated in Tristram's accidental circumcision, for instance. Having misplaced the chamber-pot, Tristram's maid Susanna instructs the young boy to "**** *** ** *** ******" (Sterne, 2009, p. 301).[3] Unfortunately, all the leaden counterweights in Tristram's house have been resumed by Uncle Toby's trusty corporal Trim, to be melted down and make cannons for their bowling-green model of the Nine Years' War. The sash window falls, and the unfortunate Tristram suffers a surprise operation. Admittedly, in this case, the bowling green acts through human agents, whose search for raw materials to upgrade the bowling green is the proximate cause of the accident. But the novel is replete with jokes the blur the boundaries between human bodies and physical objects in their environment. Tristram's mother relies on hearing the sound of Tristram's father winding the clock in order to experience sexual arousal (Sterne, 2009, p. 9). Near the end of the novel, when Uncle Toby is wooing the widow Wadman, she is curious about the groin injury he sustained at the siege of Namur. "You shall lay your finger upon the place", Uncle Toby tells her (Sterne, 2009, p. 514). She is at first surprised by this intimate suggestion, but Uncle Toby then fetches a map, and allows her to lay her finger on the geographical place where he was wounded. The map and the man become thoroughly confused. The confusion is even greater when we consider that Uncle Toby's bowling green is also the "Hobby Horse" that Tristram uses as a writing aid to describe his Uncle's character. Tristram jokes that "a man's HOBBY-HORSE is as tender a part as he has about him" (Sterne, 2009, p. 91). Like a tender body part, Uncle Toby's bowling green can cause him pain, and Tristram can use it as a creative collaborator to illuminate Toby's personality.

It is in this broad context, in which human bodies and confused with the material world around them, that the bowling green, and Tristram's imaginary writing-aids, seem to acquire agency, despite lacking some of the usual embodied and creative components that an agent would be expected to have. In its anarchic, comedic way, *Tristram Shandy* foreshadows contemporary philosophers of science, such as Jane Bennett, who argue that our conventional understanding of agency is too narrow, and that mere matter may have an "agentic power" that we overlook when we define agency in term of human intentionality and consciousness (Bennett, 2010, p. 69). Similar ideas about attributing creative agency to the materials participating in a creative act are also expressed by some computational creativity scholars, such as Bown (2015).

Tristram Shandy is not very explicit about where this "agentic power" may come from. The bowling green shifts and morphs, drawing Tristram and Uncle Toby's bodies into itself; the writing aids somehow create a connection between the writer's pen and the character's personality. These processes are joked about rather than explained. Similarly, in *Adventures of a Hackney Coach*, the Hackney Coach simply *is* sentient, with no attempt to explain this surprising fact. Aside from Coleridge, with his idea of the "one life", the only author in our sample who seriously attempts to put forward a more general theory of creative agency is Johann Wolfgang von Goethe. When Wagner creates Homunculus, he argues that matter has an innate self-organizing tendency which allows it to become creative:

> Was man an der Natur Geheimnißvolles pries,
> Das wagen wir verst ändig zu probiren,
> Und was sie sonst organisiren lie ß,
> Das lassen wir krystallisiren (Goethe, 1832, p. 105).
> [What we thought before was Nature's secret,
> That is what we now dare to experiment with.
> And what Nature once allowed to self-organize,
> We now allow to crystallise.]

Homunculus is "crystallized" out of a material that already has a natural tendency to organize itself. Even Wagner, the human scientist who is an agent in the usual sense, "lässt" ["lets"] the crystallization occur. As in the case of the Aoelian harps, where the environment itself, the wind, plays a key role in the creative process, in *Faust*, there seems to be no real difference between human agents and inert matter—everything has some kind of "agentic" or "organizing" power, and can collaborate in the creative process.

These texts present a challenging view of creative agency, which could influence the way researchers design creative robotic systems. First, if an object can be a creative agent while lacking intelligence, intention and even perception, then this could impact how we evaluate creative systems (Jordanous, 2017;

[3]"Piss out of the window"

Kantosalo, 2019). This problem of evaluation, as we have seen, is explicitly raised in Hoffmann's "Automata', when Ferdinand and Ludwig dispute the creativity and intelligence of the story's robotic systems, based on their differing attitudes about the uniqueness of human agency. Secondly, by opening up the field of how a creative agent might be embodied, these texts could inspire new designs. A creative robot could emulate Uncle Toby's bowling green: its body could be a protean landscape, that morphs in physical space to portray different information to the user. A shape-shifting landscape-robot could be well-suited for elementary education, could interact effectively with visually impaired human users, or could provide adversarial training for autonomous vehicles. The example of *The Adventures of a Hackney Coach* has already been trialled by experimental writer Ross Goodwin. His novel *1 the Road* was written by a neural network image-captioning system hooked up to a webcam that he bolted to the top of a car (Goodwin, 2018). His neural network did not meet the same embodiment and creativity criteria as Kilner's Hackney Coach (**Tables 3**, **4**), but it did make use of a vehicle's embodiment as part of the creative system.

This view of creative agency has a third challenging implication. By de-emphasizing the role of general intellect and other humanlike forms of agency, these texts foreground the creative contribution of the environment or situation. The Hackney Coach composes a highly original novel by simply recording the chance encounters thrown at it by the busy metropolis of London. Coleridge and Mörike's harps produce powerful music in collaboration with the wind. The bowling green literally *is* the environment. With their broad understanding of agency, authors such as Coleridge or Goethe found it easy to explain how the environment itself could be creative. An interesting challenge, ripe for further exploration in creative robotics, is therefore to model how the physical environment has input into the creative process, as reflected on by the "Press" (environment) variable being one of the *Four Ps* of creativity (Rhodes, 1961; Jordanous, 2016). Some useful steps forwards in this area have already been taken, e.g. Saunders' *Curious Whispers* embodied interactive creative agents (Saunders, 2012), inspired by Csikszentmihalyi's systems approach to creativity (Csikszentmihalyi, 1988); and Jon McCormack's Eden project (McCormack, 2001), an artificial ecosystem where inhabitants have a creative interactive co-evolutionary relationship with their environment. We do also however acknowledge the alternative perspective of behaviour-based robotics and related Artificial Life research, that advocates a largely representation-free interaction with the external physical environment (Brooks, 1991; Jordanous, 2020).

4.2 An Exploration of Ethics

So far we have considered these fictional examples as possible designs which scientists could evaluate or emulate. But design fictions can also serve another purpose: to help "us to explore ethical and social issues within the context of everyday life" (Dunne and Raby, 2013, p. 51). Our fictional examples reveal some of the risks posed by creative AIs, and may help scientists anticipate unintended consequences of their research. Kilner's Hackney Coach overhears the private conversations of its passengers without their knowing. Goethe's Homunculus is able to read human thoughts. On a slightly different tack, Hoffmann's Talking Turk is built to conceal its operations, and to appeal to racist stereotypes about Oriental mysticism; in these ways the Turk's creator uses it to prey upon paying customers, allowing its creator to control and profit from the public (see Falk (2021)). In each of these cases, the embodiment of the artificial agent affects how humans interact with it, with potentially disastrous results. Some of the examples are chillingly relevant today. Cars, phones and home assistants all have the capacity to record their users, and often do: What is to stop them spilling the beans, as the Hackney Coach does, particularly as text generation improves? The Talking Turk, meanwhile, offers a critical perspective on the design of chatbots and other question-answering systems. How might artificial voices or faces be designed, and how can consumers and the public be protected from subconscious manipulation?

Homunculus presents a special case. His peculiar embodiment gives him peculiar capabilities. Like the superintelligent AIs of contemporary cyberpunk novels, he seems to inhabit a world of pure information. Though he is able to hear and speak, he also emits light that grants him direct access to other characters' minds. Shortly after coming into existence, he levitates over the sleeping character Faust, and observes his dreams:

Homunculus (erstaunt)
Bedeutend! –
(*Die Phiole entschlüpft aus Wagners Händen, schwebt über Faust und beleuchtet ihn.*)
Schön umgeben!—Klar gewässer
Im dichten haine, Frau'n die sich entkleiden [. . .] (Goethe, 1832, p. 107, p. 107)
[Homunculus (astounded)
Remarkable! –
(*The phial slips out of Wagner's hands, floats over Faust and illuminates him.*)
Such beautiful scenery!—Clear water
In the shady grove, women undressing themselves [. . .]]

Through this direct mind-machine interface, Homunculus is able to peer into Faust's erotic dreams, which the sleeping professor would surely have preferred to keep private. The character Mephistopheles links Homunculus's mind-reading capabilities to his embodiment: "So klein du bist, so groß bist du Phantast" [So small you are, yet such a great fantasist.] (Goethe, 1832, p. 108). The word "Phantast" is crucially ambiguous, meaning something between "dreamer" and "novelist". There seems to be a relationship between Homunculus's tiny presence and fragility in the physical world, and his intimidating presence and creativity in the psychic world. In the end, Homunculus forsakes both his physical presence and self-identity altogether, when he fuses with Proteus in a flash of light (Goethe, 1832, p. 178). His choice of Proteus to fuse with is highly significant: Proteus is a shapeshifting Greek deity of the sea, whose body never remains in the same form for more than an instant. In some ways, Homunculus seems to foreshadow cyberpunk fantasies about uploading the mind into the cloud and living a virtual, disembodied life.

At first glance, Homunculus may seem barely credible as an AI design, but in fact he exhibits crucial properties of actually existing systems. Many contemporary AI systems exist primarily as software, and lack embodiment in much the same way as Homunculus. Likewise, Mephistopheles is quite right to suggest that Homunculus's diminutive embodiment is linked to his mind-reading abilities. Like Homunculus, devices such smartphones, home assistants, cochlear implants or Elon Musk's neuralink chip all need to be small in order to penetrate human lives or human brains. Siri and Alexa may actually reside in giant data centres, but in users' day-to-day lives they take the form of small, limbless, glowing bodies, and persistently monitor users' behaviour in order to read (or rather, model) their minds. Both Homunculus and the Hackney Coach take advantage of their embodiment to slip into human lives. In both *Adventures of a Hackney Coach* and *Faust*, the results are creative and positive: the Coach produces a brilliant satirical novel based on its secret observations, while Homunculus develops a higher form of consciousness and merges with Proteus. But the darker ethical implications are there to see, and may provoke important discussions in laboratories and design studios.

4.3 The Method of Historical Design Fiction

We conclude this discussion by reflecting on the multi-pronged approach to analysis adopted in this research. The Design Fictions approach enabled us to treat historical fiction as sources of inspiration for future creative robotics research. Our decision to use componential characterisations of creativity and embodiment has given us a multi-dimensional model by which to examine the historical texts. By considering so many different aspects of creativity and embodiment in a systematic manner, we could highlight many new and useful details to bring forward from the 18th Century to present-day attention, and we could consider examples that may not have fulfilled the requirements of the more rigid 'bifold' definition of creativity.

5 CONCLUSION

In our work we have examined several literary depictions from the Long 18th Century (c. 1650–1850) describing different kinds of embodied systems or autonomous entities capable of contributing to human creativity. Initially, the idea of examining past fictitious examples may seem curious; however the design fictions approach reveals that these examples vividly capture some of the ideals concerning creativity support and co-creativity during the time. In particular, we see an emphasis on active involvement, directed perception, experimentation and the ability to capitalize on spontaneity, with embodiment and interaction crucial to generating valuable results.

Based on our discussion of these examples, we offer four guidelines to researchers in creative AI and robotics. These are guidelines for *discovery*. They aim to widen the search space of design and research possibilities, and give researchers a finer sense of the contours of the search surface. What kinds of artificial creative agents are possible? Which are preferable?

1. **Explore unlikely embodiments:** Eighteenth-century examples invite us to widen the space of possible mechanistic co-creative partners. In *Tristram Shandy*, one creative system has a flat body of soft soil. In Coleridge and Mörike's eolian harp poems, the systems have resonant bodies of wood and gut. What new kinds of creative action might be enabled by other strange materials, body shapes and dimensions?

2. **Think of situations, not systems:** Our examples emphasize the way that systems draw on the environment to create new situations. The Hackney Coach creates a new series of connections between people and places in London, because of its imperceptibility and position in a network of human interactions. Inspired by such fictions, researchers can look beyond individual systems, and consider what new *situations* they would like to bring into being. What new environments or connections, what new patterns of interaction or behaviour might different embodiments bring about?

3. **Be aware of the disjunction between action and appearance:** How an agent appears can often conceal what it does. In *Faust*, Homunculous's tiny, fragile body conceals vast capacities—indeed, Mephistopheles suggests that his tininess actually *enables* his vast capacities. Likewise the small physical footprint of a modern agent like Alexa can conceal the large distributed system it embodies, and enable that system to penetrate people's lives. Ethical designers should consider what to conceal and what to reveal in their artificial agent's embodiment.

4. **Consider the system as a situated moral agent:** We have seen how in many of these fictions, apparently unconscious or static agents act in self-consistent and unpredictable ways, much like conscious human agents. Uncle Toby's bowling-green brings about Tristram's circumcision. Hoffmann's musical automata challenge Ferdinand and Ludwig to reconsider who is the composer of a novel tune. While it can be tempting to deny agency to mechanistic creative systems, a mechanistic system's potential consequences come more sharply into view if we consider the system itself as the one who acts.

Overall, like the example of the musical dice, the historical ideas and inspirations we highlight above can capture the imagination of modern roboticians and co-creativity scholars, and can inspire their efforts, unhindered by any potential current technical blinders. Thus, understanding how long-18th-century authors viewed mechanical/robotic creativity offers a firm foundation for building models for modern co-creative robotics. As famously observed by Sagan, "you have to know the past to understand the present."

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Authors AK and AJ provided their expertize on co-creativity and creativity evaluation and author MF acted as the literature expert for this work. The first and the second author contributed equally to the writing of the paper.

FUNDING

AK and APC funded by the Academy of Finland (decision #311090, Digital Aura).

REFERENCES

Ali, S., Moroso, T., and Breazeal, C. (2019). Can Children Learn Creativity from a Social Robot? Proceedings of the 2019 conference on Creativity and Cognition, San Diego, CA, USA, Jun 23-26, 2019 (New York, NY, USA: Association for Computing Machinery, C&C '19), 359–368. doi:10.1145/3325480.3325499

Allen, J. E., Guinn, C. I., and Horvtz, E. (1999). Mixed-initiative Interaction. *IEEE Intell. Syst.* 14, 14–23. doi:10.1109/5254.796083

Bellamy, L. (2007). "It-Narrator and Circulation: Defining a Subgenre," in *Secret Life of Things: Animals, Objects, and It-Narratives in Eighteenth-Century England*. Editor M. Blackwell (Lewisburg: Bucknell University Press), 117–146.

Bennett, J. (2010). *Vibrant Matter: A Political Ecology of Things*. Durham and London: Duke University Press.

Bleecker, J. (2009). *Design Fiction: A Short Essay on Design, Science, Fact and Fiction*. Available at: http://drbfw5wfjlxon.cloudfront.net/writing/DesignFiction_WebEdition.pdf (Accessed September 03, 2020).

Blythe, M. (2014). Research through Design Fiction. *CHI*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, '14, Toronto, Canada, Apr 26 - May 1, 2014 (New York, NY, USA: Association for Computing Machinery), 703–712. doi:10.1145/2556288.2557098

Bostrom, N. (2014). *Superintelligence: Paths, Strategies, Dangers*. (Oxford: Oxford University Press).

Bown, O. (2015). Attributing Creative agency: Are We Doing it Right?. Proceedings of the Sixth International Conference on Computational Creativity, Park City, Utah, USA, Jun 29-Jul 2, 2015 (Provo, Utah, USA: Brigham Young University), 17–22.

Brooks, R. A. (1991). Intelligence without Representation. *Artif. intelligence* 47, 139–159. doi:10.1016/0004-3702(91)90053-m

Burdick, A. (2019). Designing Futures from the inside. *J. Futures Stud.* 23, 75–92.

Butler, S. (1872). *Erewhon: or, Over The Range* (London: Trübner). Google-Books-ID: bAKBFoYiNeEC.

Chrisley, R. (2003). Embodied Artificial Intelligence. *Artif. intelligence* 149, 131–150. doi:10.1016/s0004-3702(03)00055-9

Coleridge, S., and Keach, W. (1997). *The Complete Poems Of Samuel Taylor Coleridge* (Penguin Classics).

Condillac, E. B. (1984). *Tratié des Sensations*. ([Place of publication not identified]: Fayard)

Corazza, G. E. (2016). Potential Originality and Effectiveness: The Dynamic Definition of Creativity. *Creativity Res. J.* 28, 258–267. doi:10.1080/10400419.2016.1195627

Csikszentmihalyi, M. (1988). "Society, Culture, and Person: A Systems View of Creativity," in *The Nature of Creativity: Contemporary Psychological Perspectives*. Editor R. J. Sternberg (Cambridge University Press), 325–339.

Davis, N., Hsiao, C.-P., Popova, Y., and Magerko, B. (2015). "An Enactive Model of Creativity for Computational Collaboration and Co-creation," in *In Creativity in the Digital Age (Springer)*, 109–133. doi:10.1007/978-1-4471-6681-8_7

Davis, N. (2013). "Human-computer Co-creativity: Blending Human and Computational Creativity,". Doctoral Consortium of the Ninth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE-13, Boston, Massachusetts, USA, Oct 14-18, 2013. Editors G. Sukthankar and I. Horswill (Palo Alto, California: AAAI), 9–12.

Descartes, R. (1988). *Discourse on Method and Other Writings* (London: Penguin).

Deterding, S., Hook, J., Fiebrink, R., Gillies, M., Gow, J., Akten, M., et al. (2017). Mixed-initiative Creative Interfaces. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO, USA, May 6-11, 2017, *CHI EA*, 17 (New York, NY, USA: Association for Computing Machinery), 628–635. doi:10.1145/3027063.3027072

Dreyfus, H. (1979). *What Computers Can't Do (Revised Edition): The Limits of Artificial Intelligence*. MIT Press.

Dunne, A., and Raby, F. (2013). *Speculative Everything: Design, Fiction, and Social Dreaming*. Cambridge, Massachusetts; London: The MIT Press.

Eliot, G. (1879). *Impressions Of Theophrastus Such*. London, United Kingdom: Blackwood.

Falk, M. (2021). Artificial Stupidity. *Interdiscip. Sci. Rev.* 46, 36–52. doi:10.1080/03080188.2020.1840219

Fitzgerald, T., Goel, A. K., and Thomaz, A. (2017). Human-robot Co-creativity: Task Transfer on a Spectrum of Similarity. *(ACC)*, 104–111. Proceedings of the Eight International Conference on Computational Creativity, Jun 19-23, 2017.

Gabora, L. (2005). Creative Thought as a Non Darwinian Evolutionary Process. *J. Creat. Behav.* 39, 262–283. doi:10.1002/j.2162-6057.2005.tb01261.x

Gallagher, S. (2011). *Interpretations of Embodied Cognition*.Faculty of Law, Humanities and the Arts - Papers, 1373.

Gallie, W. B. (1955). Essentially Contested Concepts. *Proc. Aristotelian Soc.* 56, 167–198.

Gemeinboeck, P., and Saunders, R. (2013). Creative Machine Performance: Computational Creativity and Robotic Art. In Proceedings of the Fourth International Conference on Computational Creativity, Jun 12-14, 2013, 215–219.

Gero, J. S. (1996). Creativity, Emergence and Evolution in Design. *Knowledge-Based Syst.* 9, 435–448. doi:10.1016/s0950-7051(96)01054-4

Gibson, W. (2016). *Neuromancer (*London: Gollancz). First edn.

Glăveanu, V. P. (2013). Rewriting the Language of Creativity: The Five A's Framework. *Rev. Gen. Psychol.* 17, 69–81. doi:10.1037/a0033646

Goethe, J. W. v. (1832). *Faust. Der Tragödie Zweiter Teil (1. Auflage)*. Stuttgart, Germany: Cotta.

Goodwin, R. (2018). *1 the Road* (Paris: Jean Boité).

Guckelsberger, C., Salge, C., and Colton, S. (2017). Addressing the "Why?" in Computational Creativity: A Non-anthropocentric, Minimal Model of Intentional Creative Agency. *Proc. ICCC.*, 128–135.

Guilford, J. P. (1950). Creativity. *Am. Psychol.* 5. doi:10.1037/h0063487

Hankins, T. L., and Silverman, R. J. (1999). *Instruments and the Imagination*, Vol. 311. Princeton University Press.

Hedges, S. A. (1978). Dice Music in the Eighteenth century. *Music Lett.* 59, 180–187. doi:10.1093/ml/59.2.180

Hennessey, B. A., and Amabile, T. M. (2010). Creativity. *Annu. Rev. Psychol.* 61, 569–598. doi:10.1146/annurev.psych.093008

Hoffman, G., and Weinberg, G. (2010). *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, 3097–3102.Shimon: an Interactive Improvisational Robotic Marimba Player

Hoffmann, E. T. A. (1957). *Poetische Werke*, Vols. 1–12 (Berlin: Walter de Gruyter).

Horvitz, E. (1999). Principles of Mixed-Initiative User Interfaces. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, May 15-20, 1999. 159–166.

Jordanous, A. (2017). Co-creativity and Perceptions of Computational Agents in Co-creativity. Proceedings of the Eighth International Conference on Computational Creativity, Atlanta, Georgia, USA, Jun 19-23, 2017.

Jordanous, A. (2012). *Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application* (Brighton, UK: Ph.D. thesis, University of Sussex).

Jordanous, A. (2016). Four PPPPerspectives on Computational Creativity in Theory and in Practice. *Connect. Sci.* 28, 194–216. doi:10.1080/09540091.2016.1151860

Jordanous, A. (2020). Intelligence without Representation: A Historical Perspective. *Systems* 8, 31. doi:10.3390/systems8030031

Jordanous, A., and Keller, B. (2016). Modelling Creativity: Identifying Key Components through a Corpus-Based Approach. *PloS one* 11, e0162959. doi:10.1371/journal.pone.0162959

Jordanous, A., and Keller, B. (2012). What Makes a Musical Improvisation Creative? *J. Interdiscip. Music Stud.* 6, 151–175.

Kantosalo, A. (2019). Human-Computer Co-creativity: Designing, Evaluating and Modelling Computational Collaborators for Poetry Writing Ph.D. thesis (Helsinki, Finland: University of Helsinki).

Kantosalo, A., and Jordanous, A. (2020). Role-based Perceptions of Computer Participants in Human-Computer Co-creativity. 7th AISB Symposium on Computational Creativity, Apr 9, 2020, London, UK (London, UK: AISB).

Kantosalo, A., and Takala, T. (2020). Five C's for Human–Computer Co-creativity—an Update on Classical Creativity Perspectives. In Proceedings of the 11th International Conference on Computational Creativity, Sep 7-11, 2020 (Coimbra, Portugal), 9–16.

Kantosalo, A., Toivanen, J. M., Xiao, P., and Toivonen, H. (2014). From Isolation to Involvement: Adapting Machine Creativity Software to Support Human-Computer Co-creation. Proceedings of the Fifth International Conference on Computational Creativity, Ljubljana, Slovenia, Jun 10-13, 2014, 1–7.

Kilner, D. (1781). *The Adventures of a Hackney-Coach*. London, United Kingdom: Kearsly, Vols. 1–2.

Kurzweil, R. (2006). *The Singularity Is Near* (London: Duckworth).
La Mettrie, J. O. d. (1996). *Machine Man and Other Writings*. Cambridge: Cambridge University Press.
Lin, Y.-T., Liu, I.-T., Jang, J.-S. R., and Wu, J.-L. (2015). Audio Musical Dice Game. *ACM Trans. Multimedia Comput. Commun. Appl.* 11, 1–24. doi:10.1145/2710015
Lin, Y., Guo, J., Chen, Y., Yao, C., and Ying, F. (2020). It Is Your Turn: Collaborative Ideation with a Co-creative Robot through Sketch. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Apr 25-30, 2020 (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3313831.3376258
Lubart, T. (2005). How Can Computers Be Partners in the Creative Process: Classification and Commentary on the Special Issue. *Int. J. Human-Computer Stud.* 63, 365–369. doi:10.1016/j.ijhcs.2005.04.002
Masamune, S. (2009). *The Ghost in the Shell 1*. illustrated edition edn (New York: Kodansha USA Publishing).
McCormack, J. (2001). Eden: An Evolutionary Sonic Ecosystem. European Conference on Artificial Life, Sep 10-14, 2001. Springer, 133–142. doi:10.1007/3-540-44811-x_13
Merleau-Ponty, M., and Smith, C. (1962). *Phenomenology of Perception* (London: Routledge).
Miranda, E. R. (2008). Autonomous Development of Singing-like Intonations by Interacting Babbling Robots. *International Computer Music Conference*. Belfast, UK).
Mörike, E. (1838). *Gedichte (1. Auflage) Stuttgart, Germany: Cotta*.
Rhodes, M. (1961). An Analysis of Creativity. *The Phi Delta Kappan* 42, 305–310.
Riskin, J. (2003). Eighteenth-century Wetware. *Representations* 83, 97–125. doi:10.1525/rep.2003.83.1.97
Riskin, J. (2016). *The Restless Clock: A History of the Centuries-Long Argument over what Makes Living Things Tick*. University of Chicago Press.
Runco, M. A., and Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Res. J.* 24, 92–96. doi:10.1080/10400419.2012.650092
Russell, S., and Allen, L. (2019). *Human Compatible: AI and the Problem of Control*. 01 edition edn.
Saunders, R., Chee, E., and Gemeinboeck, P. (2013). Evaluating Human-Robot Interaction with Embodied Creative Systems. Proceedings of the Fourth International Conference on Computational Creativity, Jun 12-14, 2013, 205–209.
Saunders, R., Gemeinboeck, P., Lombard, A., Bourke, D., and Kocaballi, A. B. (2010). Curious Whispers: An Embodied Artificial Creative System. Proceedings of the First International Conference on Computational Creativity, Jan 7-9, 2010, 100–109.
Saunders, R. (2012). Towards Autonomous Creative Systems: A Computational Approach. *Cogn. Comput.* 4, 216–225. doi:10.1007/s12559-012-9131-x
Sterling, B. (2005). *Shaping Things*. Cambridge, Mass: MIT Press.
Sterne, L. (2009). *The Life and Opinions of Tristram Shandy, Gentleman*. Clarendon Press.
Svanæs, D. (2013). Interaction Design for and with the Lived Body: Some Implications of Merleau-Ponty's Phenomenology. *ACM Trans. Computer-Human Interaction (Tochi)* 20, 1–30. doi:10.1145/2463579
Tegmark, M. (2018). *Life 3.0: Being Human in the Age of Artificial Intelligence*. (London: Penguin).
Weinberg, G., Bretan, M., Hoffman, G., and Driscoll, S. (2020). *"Be Social"—Embodied Human-Robot Musical Interactions*. Cham: Springer International Publishing, 143–187. doi:10.1007/978-3-030-38930-7_5
Weisberg, R. W. (2015). On the Usefulness of "Value" in the Definition of Creativity. *Creativity Res. J.* 27, 111–124. doi:10.1080/10400419.2015.1030320
Wiggins, G. A. (2006). Searching for Computational Creativity. *New Gener Comput.* 24, 209–222. doi:10.1007/BF03037332
Yannakakis, G. N., Liapis, A., and Alexopoulos, C. (2014). "Mixed-initiative Co-creativity," in *Proceedings of the 9th International Conference on the Foundations of Digital Games, FDG 2014. I. Bogost (Liberty of the Seas, Caribbean)*. Editors M. Mateas and T. Barnes
Ziemke, T. (2015). Czym Jest to, Co Zwiemy Ucieleśnieniem? *Avant* VI, 161–174. doi:10.26913/60202015.0112.0014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 29 July 2021
doi: 10.3389/frobt.2021.662249

The Robot is Present: Creative Approaches for Artistic Expression With Robots

Carlos Gomez Cubero[1,2], Maros Pekarik[1], Valeria Rizzo[3] and Elizabeth Jochum[1]**

[1]Research Laboratory for Art and Technology, Department of Communication and Psychology, Aalborg University, Aalborg, Denmark, [2]Human Robot Interaction Laboratory, Department of Architecture, Design & Media Technology, Aalborg University, Aalborg, Denmark, [3]Independent Artist, Aalborg, Denmark

There is growing interest in developing creative applications for robots, specifically robots that provide entertainment, companionship, or motivation. Identifying the hallmarks of human creativity and discerning how these processes might be replicated or assisted by robots remain open questions. Transdisciplinary collaborations between artists and engineers can offer insights into how robots might foster creativity for human artists and open up new pathways for designing interactive systems. This paper presents an exploratory research project centered on drawing with robots. Using an arts-led, practice-based methodology, we developed custom hardware and software tools to support collaborative drawing with an industrial robot. A team of artists and engineers collaborated over a 6-month period to investigate the creative potential of collaborative drawing with a robot. The exploratory project focused on identifying creative and collaborative processes in the visual arts, and later on developing tools and features that would allow robots to participate meaningfully in these processes. The outcomes include a custom interface for controlling and programming robot motion (EMCAR) and custom tools for replicating experimental techniques used in visual art. We report on the artistic and technical outcomes and identify key features of process-led (as opposed to outcome-led) approaches for designing collaborative and creative systems. We also consider the value of embodied and tangible interaction for artists working collaboratively with computational systems. Transdisciplinary research can help researchers uncover new approaches for designing interfaces for interacting with machines.

Keywords: artistic research, drawing, performance, dance, robot, creativity, human robot interaction, embodied interaction

OPEN ACCESS

Edited by:
Amy LaViers,
University of Illinois at Urbana-Champaign, United States

Reviewed by:
Sydney Skybetter,
Brown University, United States
Yuji Sone,
Macquarie University, Australia

****Correspondence:***
Carlos Gomez Cubero
cgcu@create.aau.dk
Elizabeth Jochum
jochum@hum.aau.dk

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *31 January 2021*
Accepted: *09 July 2021*
Published: *29 July 2021*

Citation:
Gomez Cubero C, Pekarik M, Rizzo V and Jochum E (2021) The Robot is Present: Creative Approaches for Artistic Expression With Robots. Front. Robot. AI 8:662249. doi: 10.3389/frobt.2021.662249

"Art does not reproduce the visible; rather, it makes visible."—Paul Klee[1]

[1]Creative Confession, Klee (2013).

1 INTRODUCTION

The study of the relationship between human creativity and machines has fascinated artists and engineers for centuries. The earliest machines mechanically reproduced activities associated with human creativity and artistic expression: playing musical instruments, drawing, dancing, and writing (Schaffer, 1999; Riskin and Bregović, 2017). From ancient automatons to recent applications of machine learning, artists and scholars continually explore new approaches for understanding and modelling expressions of human creativity (Herath et al., 2016; Laviers and Egerstedt, 2014). Machines designed for artistic expression function as both tools for art making and sites for creatively exploring the nature of interaction and human-machine interfaces. Identifying the hallmarks of creativity and discerning whether or how these processes can be replicated or assisted by computers or robots remain open and highly contested questions (Boden, 1994; McCormack and d'Inverno, 2012; Laviers and Egerstedt, 2014). Our interest is in exploring how robots function as creative tools and catalysts for artistic expression, and using the arts to help uncover new approaches for designing interfaces for interacting with machines. This article describes an arts-led, practice-based research investigation that explores collaborative drawing between human artists and an industrial robot. Rather than starting with a predefined research question, we conducted a series of workshops to explore how an industrial robot could be a catalyst for human creativity. Our transdisciplinary research team was comprised of artists, engineers and creative technologists who worked collaboratively over a 6-month period in a series of workshops. Together, we identified creative and collaborative processes in visual art making (namely drawing and painting) and explored how a robot could participate meaningfully in those processes. This inquiry led to the design of new tools that enabled the artist to work directly with the robot through tangible interaction in real-time. The intention of these tools was not to control or produce a specific preconceived outcome, but rather to make the robot more accessible as a tool for collaborative and creative expression. The project resulted in several tangible outcomes, including a custom interface for controlling and programming robot motion (EMCAR), custom hardware for replicating experimental techniques used in visual art, and an original human-robot dance performance titled *If/Then*. We present the outcomes of our artistic research, emphasizing the systems theory models of creativity proposed by Csikszentmihalyi and Getzels (2014) and Dahlstedt (2012). We contextualize our findings in relation to other arts-engineering collaborations as a way of thinking about the relation between creativity and robotics. We try to avoid reading creativity backwards from a finished product that traces back to an initial idea or question (Ingold, 2009), choosing instead to attend closely to the creative processes and generative movements that marked our collaboration. We reflect on the improvisational and spontaneous dimensions of the process that informed the development of an interactive system. Finally, we discuss the value of transdisciplinary research teams and arts-led approaches for designing and developing collaborative and creative interactive systems.

2 BACKGROUND

2.1 Drawing and Creativity

Drawing is a hallmark of human creativity and one of the oldest known forms of nonverbal communication. The caves in Lascaux and Pindral feature paintings from c.13000B.C., and traditional Indigenous rock art dates back even further (19,000 years). Earlier still, ephemeral drawing practices in sand are part of oral storytelling traditions by First Nations communities, where storytellers combine oral and gestural narration during storytelling rituals (Tafler, 2019). As an art form, drawing is widely recognized as a "natural extension of the visualisation of emotions, thoughts, and ideas" of human experience through the figurative use of line and materials (Wells, 2013, p.36). As an activity, drawing involves the physical act of an artist working with and through materials and tools to arrive at some poetic visual expression. We were interested in drawing as a way of exploring human-machine creativity. Drawings are produced through physical interaction with tools (a brush, charcoal, a stick, the hand, a computer mouse) and different materials (the canvas, oils or acrylic paints, sand, pixels on a computer screen). Drawing involves tactile and sensuous knowledge—what Tim Ingold calls textility—where the artist and materials engage in an artful and responsive negotiation of feeling and form. For Ingold, art works are never finished but works in progress and involve emergent processes wherein the artist uncovers possibilities by learning to "follow the materials." In *The Textility of Making* he writes, "As practitioners, the builder, the gardener, the cook, the alchemist and the painter are not so much imposing form on matter as bringing together diverse materials and combining or redirecting their flow in the anticipation of what might emerge" (Ingold, 2009, p.94). An emergent view of art making holds that the material world is not passively subservient to human designers and offers a view of creative processes as a negotiation between the artist and materials. Ingold's characterisation of the relationship between artist, tools, and materials invites parallels with Gilbert Simondon's view of how humans interact with machines (Simondon, 2016). Simondon likens humans working with technical machines to a musical conductor directing musicians in performance, where the human operator acts as a coordinator or organiser of a society of technical objects, determining the tempo of performance and managing the margins of indeterminacy inherent to machines. Ingold and Simondon's ideas about art making and human-machine interaction offer new perspectives on the relationship between human artists, machines, and creativity.

Assessing creativity in drawing usually involves an analysis of the drawing itself as evidence of some kind of poetic feeling or impulse that originates inside the artist. This limited understanding that links creativity to either an individual trait, cognitive process, or attribute of an art work has been eclipsed by systems theory models that conceive of creativity as a process between cultural (symbolic) and social forces (Csikszentmihalyi and Getzels, 2014; Csikszentmihalyi, 1998). Csikszentmihalyi observed fine art students given a drawing task and developed a systems-theory approach to describe the discovery-oriented behavior as a model for understanding creative processes. Palle Dahlstedt uses his own experiences as a music composer and improvisational performer to develop a process-based model for artistic creativity centered on the use of computational tools Dahlstedt (2012). Dalhstedt defines creative practice as an "exploration of a largely unknown space of possibilities" that can be explored through an iterative process of interaction between theoretical ideas, the attending material representations achieved through implementation, and the artist's ongoing negotiation between these two processes (Dahlstedt, 2012, p.210). The systems theory view of creativity posits that technological tools can be more than mere instruments for art making; they can embody complex behaviors and enable new lines of thought that would not otherwise be possible. At the same time, the nature of a tool sets the constraints for the exploration. If we understand drawing as something more than mere marks on a page (Walter, 1996) and instead regard it as a creative activity predicated on processes that involve human artists working with tools and materials, we can recognize drawing as a dynamic and relational process. Following Ingold, our intention is to move past the idea of an artist imposing preconceived forms on inert matter and instead consider human-machine interaction as a "looping," generative dialogue between the image in the artist's mind and the tools and materials at hand. Only then can we begin recognise how tools—be it a paintbrush, a computer or a robot—can negotiate the subtle and reciprocal relationships between the artist and material and become part of the dynamic assemblage that facilitates the creative endeavour.

2.2 Drawing Machines

Humans and tools are continually modifying each other (Stiegler, 1998; Hayles, 2012). This is true for tools developed for utilitarian practices as well as those in service of artistic expression. N. Kathrine Hayles explains the necessity of evaluating technical objects, especially digital tools, not only according to their function but as objects deeply embedded within larger social/technical processes. Following Simonodon, Hayles refers to "technical ensembles": processes and practices through which fabrication comes about, wherein the toolmaker herself is embedded in both the practice and also in a society in which the knowledge of how to make tools is preserved, transmitted, and developed (Hayles, 2012, p.88). The evolution of drawing machines, devices that through analogue or digital means engage in drawing with varying levels of human involvement, are good examples of technical ensembles. In their introduction to and edited collection of *The Machine As Artist*, Smith and Fol Leymarie present an historical overview of drawing machines and identify key conceptual frameworks and broader philosophical questions that drawing machines pose (Smith and Fol Leymarie, 2017). The history of drawing machines includes analogue, non programmable devices such as the pantograph and pendulum-driven harmonographs, to programmable automata capable of reproducing handwriting and drawing. Beginning in the 1960s, artists like Desmond Paul Henry and Harold Cohen pioneered the fields of machine and computer art. Henry's works with machine-generated effects are considered forerunners to computer graphics, and Cohen's AARON, an evolving, rule-based software program has produced numerous drawings and paintings for more than 40 years, mimicking the way that human painters work with physical materials and developing an original "style" of its own (Nake, 2012). These systems and art works intersect in compelling ways with experiments in kinetic sculpture, most notably Jean Tinguely's spectacular drawing machines. Tinguely built kinetic sculptures that produced chaotic art works according to principles of chance and unpredictability related to the mechanical designs of the machine (Salter, 2010). Tinguely's works were ultimately valued more for their sculptural properties than the aesthetic qualities of the drawings the machine produced, but they succeeded in exploring creative possibilities of technical ensembles.

Following Cohen's pioneering work with AARON, numerous HCI researchers and artists working in media art used practice-led research to explore creative potential of computers for art making. Within the field of computational creativity, artists recognize the potential of software and other computer-based tools to augment their creative processes, and, following systems theory, identify those tools, methodologies and practices that can support human creativity (Mamykina et al., 2002) (Quantrill, 2002). Michael Quantrill characterizes computers as "explorers," and uses drawing as a method of investigating human-computer integration in artistic practice without de-centering the human artist: "The idea is to use the properties of computing machines to enable forms of expression that are unique to a human-machine environment where the human is the focus, but the expression is a composite of both human and machine, in this case a computing machine environment" (Quantrill, 2002, p.218). Similarly, Oliver Bown draws on the systems theory model and Alfred Gell's notion of primary and secondary agency in his theory of computational creativity (Bown, 2012). Digital tools, like art works, can be considered secondary agents, and hint at the possibility of nonhuman agency that reveals "a gradient of agency rather than a categorical division" (Bown, 2012, p.367). While the subject of machine agency is compelling, we are more interested in investigating robots as tools for facilitating creative processes and artistic outcomes. To that end, the next section considers examples of artists working creatively and collaboratively with robots.

2.3 Robots and Art

The impact of computers on art making is well-established, but only recently have researchers begun to seriously consider the role of robots in art making. Given the connections between computer art and robotic art, it is surprising how little overlap there is in scholarship that addresses their entangled histories. Our interest in drawing robots is motivated by a broader interest in exploring how the performing and visual arts can open up new pathways for robotics and embodied interaction

(Jochum et al., 2017; Jochum and Derks, 2019). Following Madeline Gannon's work, we recognize how robots act as bridges between virtual and physical worlds ("their minds are in the virtual, but the bodies are in the physical") and as such they are not necessarily well configured or equipped for reacting to changing environments or open-ended control (Gannon, 2018, p.138). A robot's physical embodiment and material instantiation give rise to a particular set of concerns that computer art does not; embodiment has practical implications for how robots perceive and navigate the world and also for how we design systems to control and operate robots (Fdili Alaoui et al., 2015) (Wainer et al., 2006). We expand on the discussion of embodiment and interaction in **Section 6**. Many pioneering experiments in art and engineering collaborations are collected in (Salter, 2010), which includes examples from early art and technology performances (Loie Fuller's work with dance, film and lighting) and pioneering robot art works by Bill Vorn and Louis Philippe Demers (Vorn, 2016) (Demers, 2016). Within the field of robotics, Amy Laviers (Ladenheim et al., 2020), Catie Cuan (Cuan, 2021), Petra Gemeinboeck (Gemeinboeck, 2021), and Marco Donnarumma (Donnarumma, 2020) have experimented with research strategies that explore dance and other forms of corporeal expression between human and nonhuman performers. While these works vary widely in aesthetics and approach, they all share a commitment to exploring the entanglement of human-machine interaction through the staging of imaginative embodiments. Donnarumma uses the term "configuration" to denote the "performative assembly of human and nonhuman parts to create alternate forms of embodiment" (Donnarumma, 2020, p.37). These are only a handful of examples of trandisciplinary research investigations that allow artists and engineers to explore creative processes together towards new outcomes and insights.

Closer to the domain of visual arts, there are several examples of sophisticated robots drawing systems that generate drawings and dexterously work with physical materials to produce impressive drawings and paintings, including (Gülzow et al., 2018; Still and d'Inverno, 2019; Smith and Fol Leymarie, 2017; Berio et al., 2016; Santos et al., 2020). In these instances, the collaboration between artist and tools for the most part happens *via* the software, and the human artist's physical interaction with the robot is not in focus. Other artists choose to work more directly with the artist-tool-material frame, combining human artists with robot tools in real-time interaction in shared physical spaces. Sougwen Chun's collaborative drawing performances Chung (2015) and Patrick Tresset's interactive portrait drawing robot (Tresset and Deussen, 2014) are two examples of drawing robots that account for tools as technical ensembles and explore new art making practices between humans and machines. Chun's drawing performances with D.O.U.G. (Drawing Operations Unit Generation 1) began with simple mimicking gestures (similar to the pantograph), where a small robot arm reproduced Chun's physical gestures in real-time on a shared canvas. It was the exploration of the materials, especially the unintentional marks that punctured or slipped on the canvas, and Chun's improvised responses to these spontaneous and unplanned actions that render the work compelling for the artist. Similarly, Tresset's performance installations with RNP, a custom robot art and computer program for real-time portrait drawing, interrogate the role of physical presence and embodiment. The robot is programmed to draw in the artist's individual style, but the tools and materiality of the system (ballpoint pen, canvas, writing desk, robot arm controlled by servo motors, the webcam that observes the sitter and performs small animations) all direct attention to the larger socio-technical context in which the art work occurs. In both works, audiences do not merely observe a robot that makes art but are invited to observe the creative process of the technical ensemble at work, watching how human artists and robot tools continually modify and shape one another. Bruno Latour famously observed that tools are "the extension of social skills to non-humans," and these performances poetically explore the implications of tools that exhibit social and artistic agency (Casper and Latour, 2000). These art works propose different models of interaction in robotic art that account for the dynamic and temporal aspects of drawing and evidence how artists and machines can work collaboratively and creatively in ways that are not predetermined. We hope that a discussion focused on processes of becoming and collaborative creativity between artists and machines can help avoid dualistic thinking of creativity as an either/or proposition (either a machine can be creative or it cannot). We are less interested in replicating the artist's process than developing a better understanding of how robots can meaningfully participate or intervene in creative processes and designing tools that support such participation.

3 METHODS

The diverse methods used in this study reflect the transdisciplinary nature of the research team. We draw equally from the fields of arts and humanities, engineering, computer science, and human robot interaction research (HRI).

3.1 Artist-in-the-Lab/ Researchers-in-the-Atelier

Some models of creativity consider creativity to be an internal and solitary process, while others view creative processes as collaborative, improvisatory, and social. We initiated a series of workshops within the artist-in-the-lab framework. While every member of the research team has some level of artistic background, the named artist on the project, Valeria Rizzo (Rizzo), was hired to work alongside academic staff. Other members of the research team came from diverse backgrounds: Carlos Gomez (Gomez) is formally the project engineer and also a musician; Maros Pekarik (Pekarik) is a creative technologist working with interactive media and projections in live performance installations; and Elizabeth Jochum (Jochum) is a human-robot interaction researcher with formal training in theatre, dance and puppetry. The project was assisted by Andreas Kornmaaler Hansen, a graduate student in Engineering Psychology. The workshops alternated between university laboratory facilities and the artist's studio. The workshops were characterized by an exploratory, generative view of art making. The collaborative nature of our investigation acknowledges the significant role that peers play in creativity (Csikszentmihalyi, 1998). In this case, peers were not just the social environment or judges but included other lab members involved in a related research project (Jochum et al., 2020). Too often, we observe that artists are invited into research labs as creative provacateurs or instigators but rarely participate as full members of the research team. Our collaboration revealed the very concrete institutional obstacles when hiring artists to do research. It also revealed the challenges of working across disciplinary borders, especially when working with technologies that require specific knowledge or competencies (for example, programming robots). It is worth noting that these challenges are not rendered visible when artistic outcomes are presented at festivals or museums; nor are they traditionally discussed in literature. Despite the institutional and conceptual challenges of trandisciplinary research, the possibility of sharing material with multiple creative agents (e.g., other researchers) from various domains allows for more complex re-interpretations of the material. We wanted to create a rich environment across different conceptual spaces where all members of the research team could participate and contribute equally. Therefore, the project involved close, sustained collaboration where the researchers met regularly over the course of several months. The frequent exposure to other methods of working presented opportunities to participate meaningfully and learn from one another.

3.2 Workshops

The first workshop was conducted in Rizzo's studio, investigating aspects of collaboration through collaborative drawing and painting techniques and tools. We explored how these techniques could be adapted to the context of robot-human collaboration and attempted to better understand the artist's creative processes. The second workshop took place in a robotics lab, with an emphasis on trying out new techniques with the robot and observing the interaction between Rizzo and the robot. Alternating between these two workshop formats and locations, we explored working in two specific domains of creative collaboration. The primary aim was to use all the tools with similar capacity so the artist did not have to rely on engineers to get things working, and the engineers did not rely on the artist for specific instructions.

3.3 Video Cue Recall

Video cue recall (VCR) is an ethnographic method used in the social sciences and humanities (Bentley et al., 2005). Originally intended to help reduce bias in self-reporting protocols, this qualitative method aims to elicit concrete feedback from participants regarding their experiences or to conduct domain analysis. VCR has also been used in human-centered computing to gain insights into interaction behavior. We replayed the video of the entire performance of If/Then (2020) in the presence of all authors. First, Rizzo was asked to comment on her overall reactions to the performance. Then Gomez, Pekarik, and Jochum took turns posing questions and asked Rizzo to comment on specific moments in the performance. The session was recorded and transcribed using SonixAI automated transcription and reviewed and corrected by Jochum, Gomez, Rizzo and Pekarik. Jochum then reviewed the transcripts and the authors coded them according to thematic analysis. All first-person quotes from the research team that appear in this paper were obtained in this manner.

4 MATERIALS

Collaborative robots, also known as cobots, are a special class of machines. Cobots are an increasingly significant branch of industrial robots with a particular advantage over other types of industrial robots: they are designed to work in close proximity with people and are equipped with security features that adjust the force and speed of the movements to render them safe for close interaction.

4.1 UR3

The main hardware is the UR3, a cobot manufactured by Universal Robots. It is the smallest of the series, with 6° of freedom and a reach of 50 cm. This high precision robot is able to move at high speed while maintaining high levels of accuracy.

4.2 Initial Software

The initial drawing software was developed by Hinwood et al. (2018) and described in (Hinwood et al., 2018). The software was initially developed as a tool to study human robot interaction during a collaborative drawing task (Pedersen et al., 2020). The program works as follows: first one calibrates the real world coordinates of the canvas. Then raw images are entered as inputs to the software where the contours of the objects are extracted into key points, which are then translated to real-world coordinates. These coordinates are sent to the robot sequentially, so the program plots the contours that result in a drawing. The software also allows one to store animations by manually introducing a few select robot poses that the robot can execute sequentially to make the robot appear expressive and communicate with the drawer. The program uses a blocking protocol to interface with the robot, which ensures controlled speed and acceleration and making the robot safe to interact with. One drawback of this system is that the interface is blocked until each command is finished, meaning there is no possibility for real-time control. Our experience with **Section 4.2** informed the development of the new software program EMCAR **Section 4.3**.

4.3 EMCAR

To overcome the limitations of the early system, we developed the Embodied Controller for Animating Robots (EMCAR), a custom software tool for controlling an industrial robot arm that offers direct, embodied interaction for generating and programming animation sequences by manipulating the robot freely. This technique gives people with little technical knowledge the opportunity to work directly and intuitively with the robot as they would with other materials. EMCAR makes generating robot performances easier by allowing people to directly puppeteer the

FIGURE 1 | Rizzo in the HRI lab, experimenting for the first time with teleoperating the UR3 using a stylus and digital tablet. The program had several modes: mirror, follow, and replay.

robot, and also allowing for real-time teleoperation by using different inputs. For example, a Wacom digital drawing tablet allowed a person to control the motion of the robot directly through a stylus, shown in **Figure 1**. The EMCAR implementation is built on top of open-source software and made available to the community[2]. EMCAR was subsequently stress tested in the development of a human-robot dance performance described in **Section 5.2**, where a dancer interacts with a robot in two modes: one as a puppet with pre-recorded movements and the second with her body, making use of a depth camera and computer vision software that maps the dancer's body position into the robot task space. Although it was developed as a tool for artistic performance, EMCAR has potential for diverse applications beyond art.

4.3.1 Real Time Robot Interfacing

Interfacing with a robot typically involves using ROS (Robot Operating System) or the robot's individual API, using functions that block the robot until it performs a certain action. This means that the robot can be easily tele-operated, with the limitation that the robot is prevented from doing any other movement until the command is completed. This feature gives full control of different parameters, such as speed or acceleration, while ensuring millimeter precision. On the other hand, it means the robot cannot adapt to rapid changes and fluid inputs as would be expected in a performance. In other words, the system is not Real-Time Controllable. RTDE, which stands for Real Time Data Exchange, is a protocol recently implemented by Universal Robots which allows to interface the robot in real-time. The robot runs a loop with a short time of iteration from 0.4 to 2.0 s. At the same time, an external device can stream to the robot a target position, which is updated several times per second. At the end of every iteration of the robot loop, it will try to reach the last target position read. This leaves the robot free to adjust the speed and acceleration to meet the target position in time, which can be dangerous in close interactions with people and objects. Using a short iteration time (i.e., 0.8 s), it can smoothly follow any trajectory generated in real time which doesn't contain abrupt changes during the iteration time. It also introduces an intrinsic latency of this time, which is noticeable when tele-operating the robot and compromises the detail of the drawing in favor of performance time.

4.3.2 Multimodal Robot Interfacing

As a result of the different workshops described in **Section 3**, two different modes of interfacing with the robot were designed: X-Y

[2]The complete code for the EMCAR system can be found in the following repository: https://github.com/marospekarik/ur-interface.

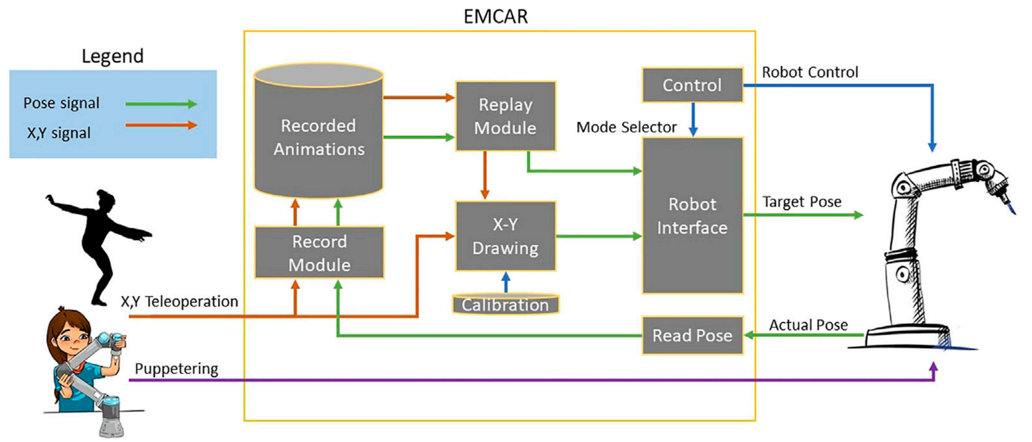


FIGURE 2 | Diagram of internal functions of EMCAR.In the left are the different inputs, X-Y position and Puppeteering. In the right is the robot interface. In the middle, inside of the orange square is the internal modules of EMCAR that create its functionalities.

Painting and Puppeteering. All this was commanded using a simple graphic user interface that allows access to all the functionalities with a simple mouse click. The diagram of the internal workflow of EMCAR is described in **Figure 2** and is explained in the following paragraphs. For the X-Y Painting mode, we designed an intuitive calibration process that allows the artist to point the four corners of a canvas on a horizontal table. This process stores the real-world coordinates where the canvas is located. After calibration, EMCAR can receive X-Y coordinates from an external device or software (for example a stylus or another sensor) and map this position into real world coordinates of the canvas and send them to the robot. As a result, the artist can use, for instance, a drawing tablet to draw together with the robot. As the X-Y input is agnostic, any other input in this format is valid. During the performance of *If/Then* the input was the X-Y coordinates of the artist on the dance floor, extracted using a depth camera with computer vision, so the robot could be controlled through the artist's movements.

For the puppeteering mode, another approach was used. Here the robot is set to "free mode," which releases the motors and allows the artist to adjust the robot manually into a desired pose. Position information for each pose during the sequence is retrieved in samples by EMCAR and stored. Each sample is composed by a time stamp and the endpoint of the end-effector. While recording, the artist can puppeteer the robot in an intuitive manner with instant feedback of what the animation will look like. This technique provides a sensuous and tactile experience for the human artist, allowing her to explore the textility of making, as a puppeteer might. The process is redolent of Ingold's description of how an artist does not impose form, but rather learns to "follow the forces and flows of materials that bring the form of the work into being" (Ingold, 2009, p.97). The embodied controller has an added advantage in that it saves an enormous amount of time and creates lifelike movements that cannot be achieved as easily with other animation methods. Once the animation is recorded, it can be replayed: EMCAR sends the desired poses in real-time, using the same time of iteration between samples used during recording. We elaborate on the implications and assumptions of embodied interaction in **Section 6**.

4.3.3 Recording and Playback

Following the puppeteering approach, X-Y Painting was also developed to record and save a drawing as an animation. As a result, the artist can store new drawings and puppetry animations in different animation banks to be used later when devising performances. These animations can be replayed by the artist (or a second operator) during a performance, moving between tele-operation on the fly and pre-recorded animations, giving more flexibility to scripted performances using both techniques.

4.4 Tablet

Graphic drawing tablets are widely used devices. They consist of a sensitive surface and a special pen that digitizes the physical strokes of a person drawing. The simplest dataset that can be obtained is the real time X-Y position of the pen, which includes information about whether the pen is touching or hovering over the tablet, and the applied pressure and angle of approach. The tablet provided a straightforward and embodied method to control the robot and produce new drawings. At the same time, it is an excellent tool for instantly obtaining X-Y positions, and was used extensively during development and troubleshooting. It quickly simulated any X-Y position generator; for example, in the performance the X-Y position was retrieved from a depth camera using motion tracking software.

4.5 End Effectors

The end effector used for drawing is a 3D printed tool, shown in **Figure 3**, that allows us to attach different drawing tools (e.g.,

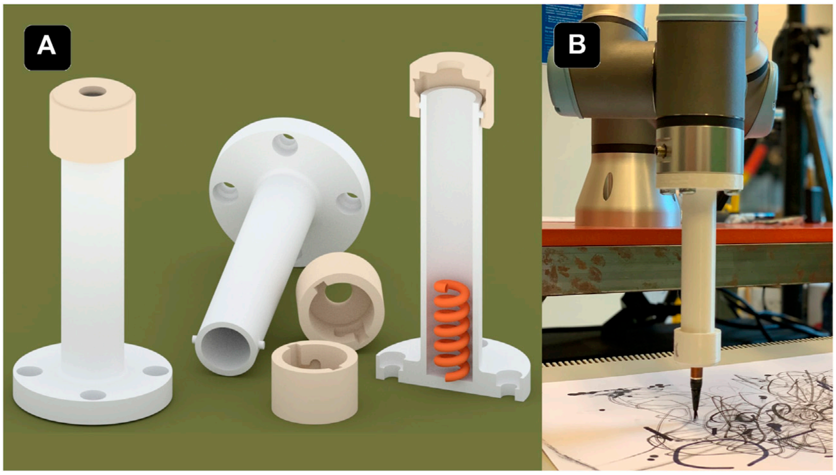


FIGURE 3 | (A) 3D representation of three of the objects used as an end effector to attach different drawing tools, such as markers or brushes. It contains a sprint in the bottom to adjust the tool. The object at the right is cut in half to display the inside. **(B)** Image of the end effector with ink brush attached to the UR3.

pens, markers and brushes) to the robot and thereby expand the artistic possibilities by allowing for a range of drawing and painting instruments. It consists of a hollow cylinder with a spring in the bottom and a cap that can be adjusted with a notch. The cap has some millimeters of headroom, allowing a small tolerance that allows for some give when drawing. The main drawing instrument used during the workshops were a ball pen with a ultra thin trace, different common white board markers with a thicker trace, a Chinese ink brush that plotted a more organic-looking stroke, and finally a professional thin white marker over a black paper that created more striking contrast and had a better finish. The strokes were unique to each instrument, and we experimented with different tools (ink brushes, pencils, markers, etc.) that each altered the appearances of the drawings. For *If/Then*, we chose to use the white marker on a black canvas, as it generates more aesthetically appealing results in a dark performance space. Markers and ball pens were used primarily in the workshops and development because of their robustness and low cost.

4.6 Projections

During workshops, we experimented with floor projections. Two ultra short throw projectors were mapped and aligned to create an interactive display on the floor. The projectors were positioned facing each other to create a seamless image by eliminating shadows. A depth camera made it possible to track the position of the performer in the space using simple computer vision techniques such as background subtraction and blob detection. The field of view of the camera was mapped with the range of the projectors that allowed for the ability to project objects at the artist's feet according to either the robot's position on the canvas, or the artist's position in the room. Projection mapping and computer vision processing were made in the TouchDesigner (Derivative, 2021), a node-based programming language for real-time interactive multimedia applications.

4.7 Limitations

It is important to acknowledge the specific limitations of both hardware and software in our project. The main limitation we experienced was the robot loop time, mentioned in **Section 4.3.1**. The RTDE protocol has an intrinsic robot loop time, where the robot tries to reach the position sent in the previous iteration. This time lasts 0.8 s, which introduces a corresponding delay. Most importantly, it overrides the data that is received between iterations, meaning that positions with less than 0.8 s are lost, introducing a "low pass filter" of the drawing strokes. Therefore, if the artist draws a zig-zag line with a frequency higher than 0.8 s, the robot won't be able to draw in time. This limitation also applies for the puppeteering function, but is less noticeable because significant changes occurring in less than 0.8 s only occur when moving the robot in a very aggressive manner.

5 RESULTS

Identifying which outcomes qualify as results can be difficult—and perhaps even paradoxical—when reporting on process-led (as opposed to product-driven) creative practice. To narrow our focus, we include a summary of the workshops wherein we identified specific artistic processes of visual art making. We then report on some of the other tangible outcomes, including a live dance performance. We also include results from the video cue recall session we conducted with the research team, as this yielded insights relevant to our discussion.

5.1 Workshop Summary

As mentioned in **Section 3.2**, two types of workshops formed the core of our research investigation. Both workshop formats were inherently distinct and designed to move the research team out of our comfort zones while also allowing space for knowledge translation between the fields. Initial workshops aimed to be a

FIGURE 4 | The research team engaged in weekly workshops at the artist's studio, engaging in collaborative drawing and painting tasks, including collaborating on a large physical canvas.

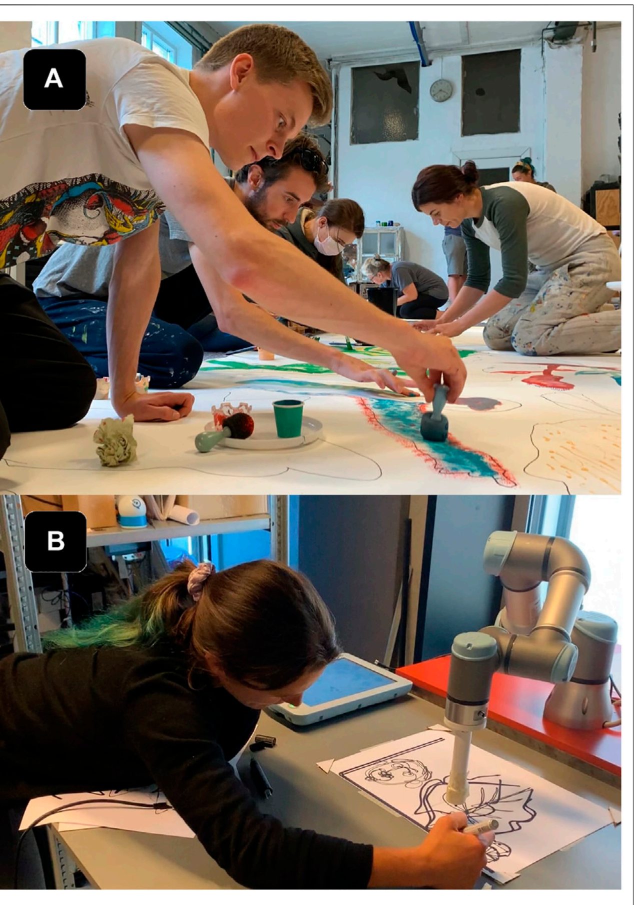


FIGURE 5 | (A) The research team engaged in several workshops at the artists' studio, exploring various painting tools and techniques with different instruments (sponges, stencils, ink pens, pipettes, etc.). **(B)** Artist (Rizzo) at the HRI lab, testing an initial drawing collaboration using EMCAR with the UR3. EMCAR allows the artist to draw simultaneously on a shared canvas in real time, as in the artist's studio. The system was an improvement over the pre-programmed drawings of the previous version of the software.

place for exchanging perspectives by introducing respective processes to one another. Rizzo led workshops on collaborative drawing and painting in her atelier. During these participatory workshops, the research team was invited to work together on a shared canvas (**Figure 4**) and to experiment with different materials and tools that are part of the visual artist's toolbox. The team spent time with various tools, trying to understand the internal process of what might encompass a drawing experience through physical interaction with the materials. The second series of workshops were conducted in the HRI lab at the university, where the research team tried to identify and translate the knowledge gained from the collaborative drawing workshops to specific methods of collaborating and co-creating with the robot (**Figure 5**). Gomez and Pekarik demonstrated UR3 capabilities through an existing software for collaborative drawing described in **Section 4.2**. As the software only reproduces pre-made drawings, the interaction was simply too narrow for Rizzo to work with. Thus, we decided to extend the design requirements. Observing and experiencing first-hand how human collaboration and co-creation developed during the workshops in Rizzo's atelier, the team understood that the real-time human's creative contribution as an input to the system and applicable responsive output from the robot might open more possibilities for creative encounters. Therefore, we prioritized the development of a system capable of real-time robot interfacing, eventually called the EMCAR tool described in **Section 4.3**.

The combination of artists' backgrounds in dance, circus performance, theatre, puppetry, music and interactive art influenced the next development stage. Consequently, we expanded the activity outside the drawing format, which led us to devise a performance for live audiences. We implemented tele-operation features controlled with the stylus (described in **Section 4.3.2** and seen in **Figure 6**), which gave Rizzo a sense of the ability to control the robot and produce drawings. However, the full control over the system had adverse effects on the aspects of co-creation. Rizzo was less interested in having precise control over the robot to intentionally make marks on the page, and more interested in interfacing with the robot intuitively, the way she worked with other tools and other artists in the first workshops. What is important to note is the transition from visual art to the study of physical motion. The experiences of co-creation between human-partners through tools on a shared canvas opened up a line of inquiry that we had not fully considered: the movement of the artist and the negotiation between the human artist and the robot was essential for creating an experience of collaboration. From the second workshops, Rizzo expressed an interest in moving together with the robot to produce a drawing, and thus our focus shifted to creating an interface that allowed Rizzo to work in a more physical, though less deterministic, way with the system. We focused on ways to translate Rizzo's motions to the robot's body, drawing on corporeal empathy and making explicit the connection between human-robot-tool (Fdili Alaoui et al., 2015; Sheets-Johnstone, 2011).

FIGURE 6 | The research team in a workshop at the HRI lab, experimenting with an early prototype of teleoperating the UR3 using a stylus and digital tablet.

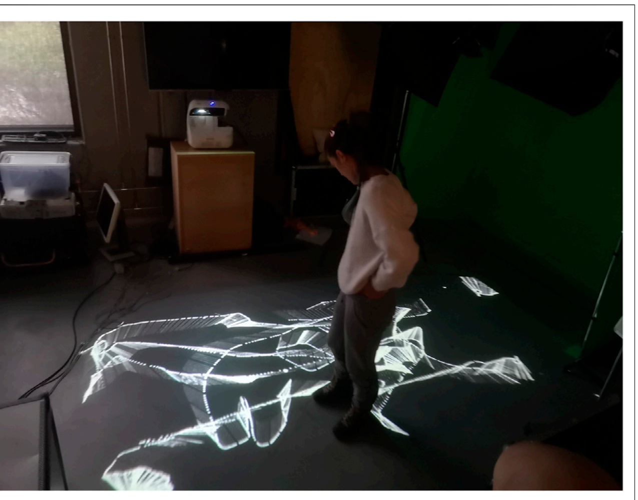


FIGURE 7 | Interactive projections on the floor track the motion of the Movement trajectory traces represented with a projection on the ground.

Once we translated dance moves to the robot's motion, we began to conceptualize the entire physical space as a canvas. The artist's position and movement in the space were highlighted by projected animations on the floor described in **Section 5.2.3** and seen in **Figure 7**, mirroring the stroke on the physical canvas drawn by the robot. It was important to work with a technology that would allow Rizzo to have freedom of movement without any unencumbered body movement. Pekarik and Rizzo had previously collaborated on an experimental performance involving physiological sensors, motion tracking and dynamic projection mapping[3], and Jochum had previously developed an improvised robot-dance performance (Jochum and Derks, 2019). Given our shared background in performance technologies, we decided to work with the combination of depth camera and projection mapping techniques to make the material more tangible for the performers and the audience.

The last workshops in our process focused mainly on devising the performance of *If/Then*, incorporating sound, lights, live-feed video cameras, and visuals that complement the performer's actions, shown in **Figure 8**. The choices concerning the narrative are presented in **Section 5.2.1**. To further convey the narrative aspects of the performance, the EMCAR tool was extended with a puppeteering mode described in **Section 4.3.2**. The immediate recording and replaying of the animations offered a high level of physicality and embodied interaction which allowed the team to work together to intuitively explore possible motions, illustrated in **Figure 9**.

FIGURE 8 | The research team iterating the design for the projection animations and performer tracking during a workshop at the RELATE lab. On the screen is projected a real-time feed of the performer and animations captured from a video camera mounted above.

5.2 Performance

The outcome of the workshops resulted in an original performance staged three times at the Danish National Museum of Science and Technology. The performance was conceived as a complementary program for the interactive drawing installation that ran during the day. The performance was not meant to be a final showing, but rather a public demonstration of a work-in-progress. The duration was around 15 min, and was performed on the half hour, with three

[3]Video of the In-Pulse performance at https://youtu.be/0nMKvoos6TQ

FIGURE 9 | Rizzo works with the robot during a puppetry workshop at the RELATE lab. Using physical manipulation, EMCAR allowed the artist to work directly with the robotic arm to choreograph, record and playback animation sequences.

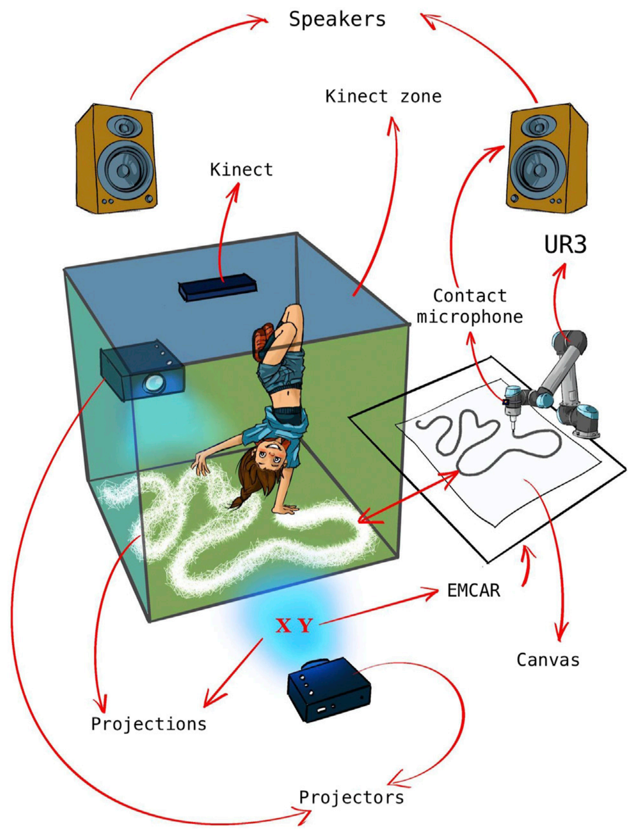


FIGURE 10 | Overview of the final of the system and implementation details for the live performance. A projector mounted on the ceiling projected real-time video capture from the performance from four different angels (including a camera mounted directly on the robot) on one wall in the gallery space. Two floor mounted projectors using sensor data from the cameras and projected lines that corresponded to the movement patterns of the dancer with lines on the floor.

performances in all.[4] In **Figure 10** is presented the setup used in the performance. The initial agreement with the museum was to exhibit an interactive drawing robot, similar to the one described in (Pedersen et al., 2020), although this time with a smaller robot (UR3 instead of UR10). The process of generating a public performance in a space that was not designed for performance, and with a producing partner with no prior experience with live performance events, meant that the performance came into being because of circumstances surrounding the installation rather than a specific idea or pre-formulated script. In this way, it echoes the process-oriented view of artmaking and technical ensembles view described above. The narrative and dramaturgy of the performance was a direct outcome of the practical necessity of running an installation during the day that would seamlessly transition into a performance without disrupting the museum. The public presentation, although documented and recorded, was never considered a finished product, but a material expression of the investigation of the limits and possibilities of real-time human-machine interaction, a process of becoming as described by Ingold.

5.2.1 Narrative

The ability to experiment physically with the robot and work with the projection system inspired Pekarik and Rizzo to form a narrative around the already existing research activity conducted with the robot. Therefore, Rizzo's role as a research assistant was integrated into the final narrative in favour of our research continuation to explore her partnership with the robot further. The narrative followed the journey of a team of researchers undertaking a routine examination of a robotic arm tasked with a routine drawing operation (standard procedure during our investigative practice). Arriving at the workplace, the researchers find the robot stuck with an unexpected drawing output on the desk, shown in **Figure 11**. The anomaly indicates a possible bug in the robot's system that could have occurred overnight. The operators (members of the research team) restart the system and perform check-ups that confirm the robot is ready to resume work. However, during these procedures, the robot becomes distracted and breaks away from the task to look around. The researcher tasked with supervising the robot suddenly notices the strange behavior, and the robot ceases the predefined drawing task and begins to create a new drawing that is mapped to the researcher's position in space. Taken by surprise, the researcher responds with curiosity, and subsequently engages in a movement exploration to investigate the mappings of her motions in space to the robot's movement on the canvas, and the corresponding light drawings that are projected in real time on the floor. The exploration intensifies until the robot suddenly drifts off the canvas, inadvertently scattering items across the table. The performance ends with researcher and robot facing one another in tableaux.

5.2.2 Puppetry

During the workshops, the performer recorded various robot animations with EMCAR's recording functionality described in

[4]Video of the full performance *If/Then* available at https://vimeo.com/491681339

FIGURE 11 | The team performs *If/Then* performance at the museum. Rizzo works with the robot and executing a choreographed sequence in front of the audience. She controls the robot in "free mode" to get a drawing out of the drawing area.

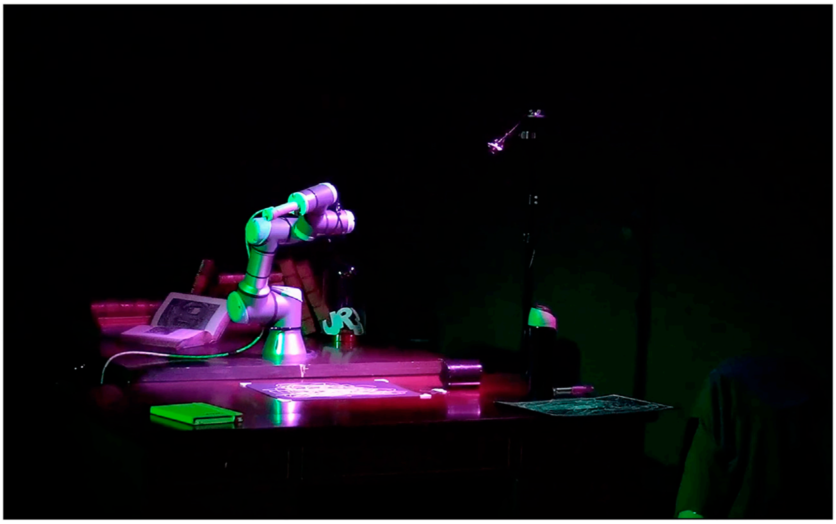


FIGURE 12 | The robot after performing animation of hitting a cup from the table during the *If/Then* performance at the museum. EMCAR was used to record the animations before the performance and replay them in the cue moments.

Section 4.3.3. Two types of animations were recorded, using puppeteering mode and drawing mode described in **Section 4.3.2**. In puppeteering mode, the artist recorded animations by physically manipulating the robot to the desired sequence. This made it possible to create expressive human-like animations for the robot like the "hitting the cup" shown in **Figure 12**. The drawing mode allowed the artist to recreate a real-time drawing process captured on the tablet, and was used for the drawing loops on the canvas. No attempt was made to hide or mask the operation of the robot: during the performance, animations were cued and executed by team members seated onstage and in full view of the audience. The performance combined a mix of pre-recorded and live tele-operated actions that, together with the performer's improvisations, meant a unique performance (and drawing) each time.

5.2.3 Interactive Projections

The system included three projectors, one that projected composite, real-time images from four unique camera angles, and two that projected on the floor overlaying each other where the artist stood, creating an interactive "screen" that was mapped with the tracking camera data and the robot canvas. The floor then became a real-time visual feedback of the movements of the robot showing the path that the robot was following, shown in **Figure 13**. This setup allowed moments for the artist to break "eye contact" with the robot, shifting her attention and allowing more general freedom of movement in the performance without breaking the dialogue with the robot. Together, the multiple projections created an interactive environment that created an immersive

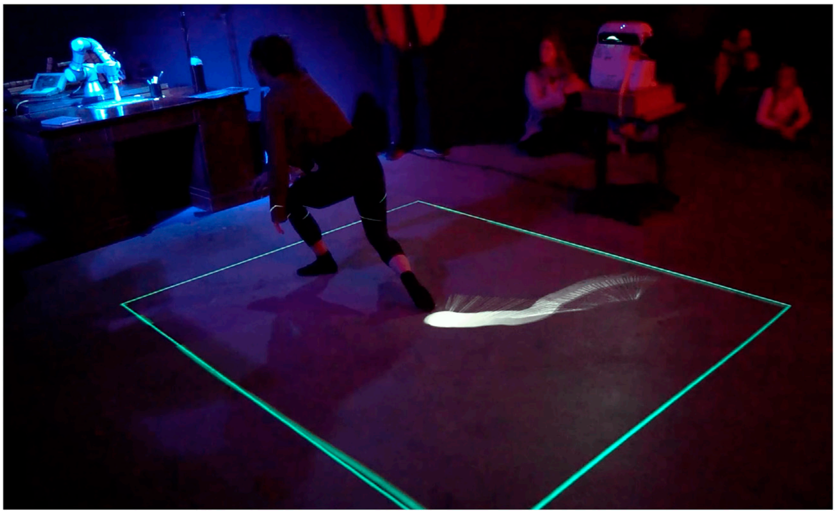


FIGURE 13 | Projectors facing each-other, creating an interactive display. The display is mapped with the tracking camera and the robot canvas. The floor becomes visual feedback of the artist's movements controlling the robot.

space for both the performer, the operators, and the audiences, inviting the possibility for multiple perspectives on the scene as shown in **Figure 14**.

5.2.4 Sound and Movement

The sound for the performance was generated using a contact microphone attached to the robot. The natural sound of the robot motors are fed through a filter that produces a grating, crackly sound that modulates with the movement of the robot. When the robot draws, the noise from the sound of pen making contact with the paper is amplified, calling attention to the acoustic properties of drawing. The sound is sent through two powerful speakers positioned overhead, resulting in a loud and uncanny soundscape that amplifies the presence of the robot. In addition, we used a Korg Synthesizer Monotron to add two special effects, delay and distortion, to enhance the sound at key moments during the performance. A keyboard was used to add a simple melody in at the climax of the performance. The sound was performed live by Gomez, the team engineer who is also a musician. From the outset, we knew that we wanted to explore the sounds of the robot, rather than a separate score. We conducted some early tests in the lab with contact microphones on the robot, that produced sounds that were passed through filters to generate interesting effects. This approach was inspired by previous work with contact microphones on robots and also work by Schacher and Wei (2019) that mapped brush gestures in Chinese calligraphy with sounds processes during a live performance with two performers on a shared canvas (Schacher and Wei, 2019). Our explorations revealed that we can use the amplified signal from contact microphones to achieve sonification of movement without any synthesis techniques. The performer's body movements would facilitate the creation of mechanic sounds, strengthening the robot's presence and also making clear to the audience the connection between the performer's movements and the robot's motions that produced the drawing.

FIGURE 14 | The artist performs *If/Then* performance at the museum. The four channel projections on one wall of the gallery were comprised of four real time cameras that alternated throughout, giving the audience a chance to observe the performance from various perspectives. The video channels were VJ'ed live by Pekarik. Jochum designed and operated the lights, and Gomez performed live mixing of the real-time sound score.

5.3 Video Cue Recall

We conducted a video cue recall session following the performance. The topics of the conversation were not limited to the performance. Rather, the video was used as a baseline for generating a discussion about various aspects of the collaboration, including the initial research stages. Thus, the video session evolved into a semi-structured interview. We attempted to steer the conversation back to the performance, which played continually on a loop throughout the interview. We sometimes paused the playback, either slowing down or speeding up sections to focus on specific moments and review them. Loosely following the principles of thematic analysis, we identified three themes from the session: Interaction as Game, Improvisation as Dialogue, and Embodiment. Considerable discussion was given to possibilities for revising the performance or expanding it in other domains. We summarize these three themes with the view that they might be useful for facilitating creative expression for other artist/researchers interested in human-robot collaboration.

5.3.1 Interaction as Game

Reviewing the performance, Rizzo described her interaction with the robot both during the performance and during the workshops. Gomez, Pekarik and Jochum were impressed by Rizzo's candor when expressing her lack of enthusiasm for the original drawing software introduced during Workshop #1 (Pedersen et al., 2020). She expressed her dissatisfaction not only with the artistic quality of the drawings, but primarily because she didn't see any possibility for real interaction with the system: "The very first drawing program involving robots that you presented to me was totally uninteresting. I was really in doubt about how it could be useful to me as an artist. This machine that is making (and I'm sorry to say it) but very stupid, simple and ugly drawings. And I was like, well, "How useful is it for me?" and "How entertaining is it for people looking at it?" "How interesting is it for artists?"" Rizzo elaborated that while the interaction with the first software program describe in **Section 4.2** was centered on a simple guessing game (image recognition), the EMCAR tool allowed her to develop a more complex game that was open-ended and playful, thereby giving her more possibilities to explore as an artist: "If a person is looking at a robot drawing an elephant, people might say, 'Ah, it kind of looks like an elephant.' Or 'Yeah, it kind of looks like an animal.' And the game finishes there. To attract interest from people, you have to start a game. And the game finishes the moment people realize what the robot is drawing. There is nothing else to guess, there's nothing else to see, or to imagine. You don't want to discover more. So I think that we should leave the game a bit more open and unclear, open to investigation and imagination, so the interest of the people stays high, at least for a longer time. And with this one, I think it's easier to keep it open for longer time. And also, it's more interesting because every single person will see different things and will be inspired in different ways." Rizzo also referred to the entire performance situation as a type of game, this time invoking the suspension of disbelief that is intrinsic to all theatrical productions: "There was this game (that obviously we know was fake), but there was this game that it was looking like the machine was still expressing itself, even though it was following me. And in that way, I feel inspired and I feel that I could work with it."

5.3.2 Improvisation as Dialogue

When asked about generating improvisation during the performance, Rizzo referred to improvisation as a kind of dialogue centered on nonverbal communication: "For improvisation as a solo performer, I believe you follow an inner path. So there is a dialogue inside of you. And for the dialogue you just decide who you are talking with. And sometimes you're talking with your audience, sometimes you are just talking with yourself, in this case, sometimes I was "talking" with the robot. But you decide where to go. You decide which language you want to talk. You decide if you want to stay silent. You decide if you want to laugh or if you want to scream. So as a solo performer, life is easier. If you are in a group, you constantly have to deal with the others, to consider the others, look at the others, listen to the others. That is why improvisation in a group can be so difficult. Sometimes you lose inner concentration, you lose the inner peace that allows you to be clear in your intentions, in your dialogue intentions, what you want to say. Because a performance is nothing but a speech. You are saying something. And if you are in a group it is very difficult. That is why normally group improvisation, they have kind of boxes, or boundaries, or pre-decided limits like, if all of the sudden one person is doing this, then everybody follows that."

5.3.3 Embodiment

When asked about working with EMCAR, Rizzo described the software presented in the early workshops, which used pre-loaded images and didn't allow for the ability control the robot in real-time. "I remember when I arrived in the lab, and you presented me with the robot that drew from files that you physically put in the computer, and I saw how the robot just recreated the drawing in a very precise and ugly way. I thought, "You know that we have printing machines, right?" Obviously I'm not trying to take down all the work that was behind it, but as a person that doesn't know anything about programming, I believe that others would also think the same thing that I did." Without being prompted, Rizzo compared the first software program to her experience working with EMCAR, beginning with the tablet: "So now you have a machine that I think any other person might find interesting. Because, again, we have inputs that we can put through the tablet [..] so it's not drawing by itself, but it's drawing through the hand of another person, and creating amazing landscapes for performers." Rizzo was also excited about the expanded possibilities offered by the puppeteering function: "Even a robot that is a performer itself, because of the puppetry movement, that is actually an awesome thing that I really got inspired by. I thought, "Oh my God, you could do so many things with that!" And again, you also showed that it was very interesting for people to try it out. Through the puppetry movements, people also realize better how to express themselves. Like, if I want to express anger, what should I do in order to be simple? If I want to express sadness, what should I do? So people also realize more how to express those feelings through clear movements made by a machine that is absolutely without personality—no feeling, no facial expression—just a machine. The robot doesn't even have a fake face, it doesn't even have eyes. But still, with the right movements, that machine is alive and you turn it into a sad machine, an angry machine, a happy machine. For me, on the inside, it was very important because I was guided by the robot, but I also felt that I was guiding the robot."

6 DISCUSSION

Our project began with a custom software tool that allowed for collaborative drawing with a robot. The first drawing tool was rather naive, and the initial work did not include any collaboration with artists. We assembled and engaged a transdisciplinary team of researchers, including a professional artist trained in classical painting, dance, and circus performance, to explore the creative potential of human-robot interaction. We did not have a predefined goal or research question at the outset. Instead, we proceeded from an arts-led, practice-based research perspective to explore possibilities for human-robot creativity. Our motivation was to identify creative processes associated with visual art and understand where and how a robot might meaningfully intervene in these processes to support human creativity. When we began our project, we had no idea that we would end up making a dance performance nor did we have an idea of what tools would be necessary to make that performance realizable. This is evidence for how arts-led, practice based research can facilitate discovery-oriented behavior through discovered-problem situations. (Csikszentmihalyi and Getzels, 2014). The tools we developed were those that the artist needed, born out of exploratory practice and the textility of making and not from some preconceived idea that originated in the engineer's or the artist's head. The focus on process-led discovery also called attention to the dynamic relations between artistic team, tools, and materials, and eventually the audience. In the systems theory model, creativity is not an attribute inherent to a product or artefact, but depends on the effect it is able to produce in others: "What we call creativity, then, is a phenomenon that is constructed through an interaction between producer and audience" (Csikszentmihalyi, 1998, p.314). On the most basic level, our project demonstrated how the robot as an interactive system came to be regarded as creative because the performer shifted in her response and reaction to it through a constructivist approach. Through the tools and performance, Rizzo gradually came to regard the system as interactive, as something that she could actually work with. Rizzo's characterisation of both the performance and the interaction with the robot as improvisatory and dialogic echoes Ingold's concept of creativity as a becoming process that brings together diverse materials by "combining or redirecting their flow in the anticipation of what might emerge" (Ingold, 2009, p.94). The dialogue that emerged was not only between the performer and the robot, but involved the performance environment (projected light that animated the floor in response to the performer's movements), materials (brushes and ink, canvas), sound, and the audience. Like Simondon's technical ensemble, Rizzo became a kind of conductor during the performance, coordinating the action and network of tools and materials as well as the activities of the other members of the artistic team seated onstage at their computers.

Observing her performance with robot, Rizzo described her interaction with the robot as a kind of game. The strategy of game as a concept for designing has been studied in the context of interactive media art, (Kluszczynski, 2010), but has yet to be taken up in human-robot interaction. According to Kluszczynski, the Strategy of Game organizes events and outcomes that emerge from the interaction itself. A basic characteristic of this strategy involves a specific task to be performed, where each participant has access to the rules and tools of the game and a certain amount of space. The strategy of game differs from games because it draws the attention of participants "not only toward the tasks that are outlined, but also toward the interaction's course, its architecture, relations between the game's structure and its properties, and also the other discourses included in the event." Art works that utilize the strategy of game "place in the discursive opposition not only the player and the game, but also the process of playing, in this way gaining the possibility to make all these aspects of the game and the game world as understood generally debatable" (Kluszczynski, 2010, p.8). Another feature of this strategy is that it allows for the possibility to approach metadiscursive issues that are not directly connected with the game or outcome, thereby enabling the artwork/interaction to develop discourses within its own structure that are critical toward the game/task. One can imagine approaching interaction design and interfaces for human-robot interaction that allow for this kind of critical engagement. The result could be an interface that aims at intuitive, natural interaction while making clear the underlying logic and limitations at work in the system.

Artists' experimentation with conceptual and material representation plays an integral part in the creative process (Dahlstedt, 2012). The artist can explore more intuitively the possibilities of what a robot can do when the system offers interaction in a natural way that echoes her process of making, not only the outcome or product. Embodied interaction is an interaction with technology that offers an opportunity to interact with the system naturally. As Dahlstedt (2012) points out, new ideas are more likely to emerge from the iterative process where the artist is directly engaged in a dialogue between conceptual and current material manifestation. The important aspect is that the material offers the possibility for this type of this communication—like a sculptor working with marble. The advantage of a system that uses embodied interaction is that the artist is empowered to refine possible conceptual and material spaces with more ease. Working through the material's resistance can challenge the artist's desire to shape the form. As mentioned in Nake (2012), artistic expression requires that the artist finds creative ways to work with or through resistance of the material, in order to shape it. Physically, materials occupy a spectrum of resistance. According to Dahlstedt (2012), tools offers navigation in the limitless space of intrinsic material possibilities, but only along the paths that the tools provide. If navigating those paths can become intuitive, the process of exploration is accelerated, which results in an artist's expression. In this sense, our program stands apart from algorithmic and digital art. Even though we were working with software, the nature of the system, through its embodied interaction capabilities, influenced how we navigate that space of possibilities. For instance, programming robots using embodied control allows for greater accessibility that makes working with robots more accessible for people without engineering or computer science backgrounds. We took inspiration from puppetry, where traditional puppeteers enjoy immediate feedback by working directly with the material/puppet. This creative process typically depends on immediate response and force feedback of the animated object, which help the artist to design choreography intuitively and create expressive movements/ animation. For artists not used to working with technology, working without this direct feedback can be challenging. The ability to control the robot with her body or through a stylus gave Rizzo a completely new perspective on the machine: "It's a completely different machine. Now, it's a colleague, it's a pal that I would like to work with. I'm looking forward to work again with it." The importance of embodied computing and its relevance for meaning making and perception is well documented (Wainer et al., 2006; Sheets-Johnstone, 2011; Fdili Alaoui et al., 2015).

Recent scholarship in HCI, informed by Disability Studies and critical feminist scholarship, has highlighted the ways in which the conventional approaches to design for embodied interaction are highly problematic (Giaccardi and Karana, 2015; Shildrick, 2013). As Katta Spiel notes in (Spiel, 2021), "bodies and how we design for them are products of social norms," and these norms contain dangerous adverse consequences for bodies and people that do not fit readily inside these normative categories. Much of HRI and literature on embodied interaction equate being human with white, male, non-disabled bodies. The implicit Western male whiteness contained in the conceptualisations and artefacts in the field of embodied computing are more than mere blindspots, they materialize and encode bias and do not account of the experiential differences in lived embodiments of women, BIPOC or people with disabilities. The result is that practices in the field of embodied computing fail to account for the "axes of oppression" that reify certain forms of power, rendering it all but impossible to rethink or design for bodies outside of normative categories. Unfortunately, critical inquiries like Spiel's do not feature prominently enough in HCI or HRI research, although there are promising signs that this practice is beginning to change. Design for embodied interaction that allows for plurality and difference of human embodiments can and should be considered when designing embodied controllers or devices. In our project, we focused primarily on developing tools for Rizzo that would not require programming skills or understanding the underlying logics of the system. Rizzo is a non-disabled dancer with decades of training in somatic and dance practices. We were attentive to the lived, bodily experience of the artist working with the tools and the difference in how she encountered tools in her atelier versus the tools in the lab. Our intention was not to encumber Rizzo with gadgets or tools, but to provide an embodied experience that was reminiscent of the tools and the way she worked with those tools in her own visual art practice. The initial experiments with the drawing tablet and stylus were familiar to Rizzo from her work with computer drawing tools as a children's book illustrator. However, the drawing technique for controlling the robot motion did not do much to inspire her. It wasn't until Rizzo was presented with the motion tracking technology and moving projections on the floor, which allowed her to directly observe the link between her physical movements and the movement of the robot, that she began to feel inspired to work creatively and empathically with the robot.

Gemeinboek (after Dautenhahn) problematizes the notion of corporeal empathy and embodied interaction for designers: how does one design for embodied interaction when there is no such thing as "natural interaction"? (Gemeinboeck, 2021) As shown by Fdili Alaoui et al. (2015) and Gannon (2018), human-centered interfaces can enhance, augment, and expand human capabilities through bodily extensions or worn prosthesis. Typically these devices rely on sensors or other wearable controllers that control or direct the movement of the robot, usually through remote tele-operation. Such devices can be read as prosthesis. Disability studies scholar Margrit Shildrick has advanced critical perspectives that link technologies and devices with affective experiences and subjectivity. Shildrick's notion of embodiment and embodied interaction explores the "affective significances of prosthesis and devices that transform the body, demonstrating how corporeal transformations can work to undo the coventional limits of the embodied self" (Shildrick, 2013). She identifies in prosthetic devices a potential for a "celebratory reimagining of the multiple possibilities of corporeal extensiveness" (Shildrick, 2013, p.271). While the tracking technology we experimented cannot be called a prosthesis, the fact that Rizzo was able to control the robot and produce two sets of drawings—one on the canvas of the floor through projected light, and the other through the robot and the canvas on the desk, we can read the entire system as a kind of technical ensemble, or a type of prosthesis that expanded the conventional limits of Rizzo's body and triggered her imagination. The convergence of artist-tool-material-space brought about a new corporeal configuration that begin to make possible a creative re-imagining of alternate forms of embodiment and artistic expression (Donnarumma, 2020). It is also interesting to note how the experience of working collaboratively and creatively with the artist impacted the perspectives of the other members of the research team in ways we could not have imagined beforehand. For example, reflecting on the workshops, Gomez (an engineer) commented that the entire experience changed his perspective on how he would approach research problems in the future. For example, his next project involves using a CNC machine to carve mortar for facades. He remarked that before beginning development on that project, he would begin by exploring the technique by hand, in order to gain an embodied understanding of working with and through the materials. Our arts-led, practice-based investigation reconfirmed the necessity of tactile and sensuous exploration and knowledge of materials, knowledge that has long been considered tangential to cognitive theories of creativity, but deeply entangled with creative artistic practice. We learned that embodied exploration of material was not only important for the artist (Rizzo) when working with robots, but also for programmer (Pekarik) and engineer (Gomez) responsible for designing the interactive systems. During the weekly drawing and painting workshops, the research team experimented together using different tools and collaborating on a shared canvas. The sustained interaction allowed the partners to delve more deeply into each other's world and material practice, providing us with an embodied understanding of artistic processes and tools that we would not otherwise have access to. Positioning the canvas on the ground and collaborating together on a shared canvas both defamiliarized the activity of drawing and invited another way of knowing and relating to materials and to one another.

Thinking through the material is key when designing tools or systems. Tools are, of course, extensions of the artist, although the artist does not necessarily need to be able to produce the tool in order to utilise it. Engineers, on the other hand, are specialized in creating tools that allow others to explore the material creatively. The sustained interaction among the members of the research team generated a bond, that through iteration grew stronger and resulted in embodied knowledge exchange and appreciation of different perspectives. The different workshops helped to generate this bond and to find common ground where the desires and expectations of the artist and engineers met from functional, reliable and safe perspectives. Reflecting on the co-creative aspect, Pekarik expressed that understanding the intrinsic motivation of the drawing activity as a communication process between artist and the material helped him to prioritize design decisions towards embodiment qualities. The authors all agree that this close collaboration enlightened the best practical possibilities and positively influenced the research outcomes. Although the process resulted in new software and hardware tools for artists to work with, to regard these tools as creative in themselves would be shortsighted. These creative outcomes are not finished products, but artefacts that open up new possibilities for creative exploration across new topologies. Rather than products that signal creative outcomes, they function as material evidence for creative processes. We plan to continue working with these tools and processes to develop more diverse tools for the artist-robot team to explore, both in the laboratory and in the atelier. Current research exploring expressive robot animations by Pakrasi et al. (2018) and real-time interaction with "live" algorithms in performance by Blackwell et al. (2012) indicate possible future directions. Through our investigation, we widened our own conceptual models of human-machine interaction and co-creation. Art involves generative processes that require negotiation and interaction with physical materials and tools for art making. Artistic and creative processes are not confined to human-tool interaction, producer-audience relations, or product-audience judgements. Artistic creativity is capacious: it extends to the environment and involves an entire network of physical and digital objects, organic and inorganic, artificial and natural, entangled in a field of relations that is continually shifting, recompiling, and interweaving between physical and virtual spaces, through planned and unplanned actions. If live performance is where the planned and the unexpected meet, we can imagine no better site for creatively exploring new possibility spaces for robotics and human-robot interaction.

7 CONCLUSION

Typically, problems in robotics take the form of presented problem situations, where the problem and tools for solving the problem are known at the outset. Our exploratory, trandiscplinary research began with a different intention: utilizing creative methods, we generated discovered-problem situations to generate new ideas and approaches for designing interactive systems and human-machine interaction. Rather than focusing on a robot that could produce artistic outcomes, we focused on drawing as an activity that could help us explore more deeply "the itinerant, improvisatory and rhythmic qualities of making" (Ingold, 2009, p.99). Drawing is intrinsically dynamic and temporal, and can be understood as a process of becoming, rather than being: "You cannot be a mountain, or a buzzard soaring in the sky, or a tree in the forest. But you can *become* one, by aligning your own movements and gestures with those of the thing you wish to draw. [. . .] As with the mountain path, the buzzard's flight or the tree root, the drawn line does not connect predetermined points in sequence but 'launches forth' from its tip, leaving a trail behind. [. . .] It has no end point: one can never tell when a drawing is finished" (Ingold, 2009, p.99). Our project demonstrates the possibilities of reimagining human-machine collaboration and technical ensembles. We found strong links between artistic creativity and discovered problem-solving processes. Iterating and developing ideas in an open-ended (as opposed to predefined) manner altered, evolved, and expanded the outcome of our creative process in ways that we could not have anticipated. This process was reminiscent of what Ingold calls "looping" - the processes of an artist working directly with tools and materials in a dialogic manner. The concept of dialogue emerged as a salient feature for both improvisation in performance, and the human-robot interaction during the collaborative drawing sessions. Transdisciplinary research facilitates creative processes between humans and machines, allowing the interactions to take shape with and through materials in dynamic and collaborative encounters.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/marospekarik/ur-interface.

AUTHOR CONTRIBUTIONS

CC, MP, VR, and EJ are co-authors on this paper. The funding for the project was secured by EJ. The artistic concept, development, and implementation was a joint collaboration between all authors, and CC and MP led the technical development and design of EMCAR. All authors contributed to the artistic and academic aspects of the project, with support from additional team members named in Acknowledgements.

FUNDING

Funding for the project was provided by Helsefonden award number: 20-B-0418. With additional support provided by the Research Laboratory for Art and Technology (RELATE) and the Human Robot Interaction Lab at Aalborg University.

ACKNOWLEDGMENTS

The authors would like thank Peter Bjerregaard, Torkil Anderson, and Jacob Thorek Jensen at Denmark's National Museum of Science and Technology for hosting the performance. We would also like to thank Sebastian Bülow and Thomas Kristensen from AAU Media Lab for their creative input and material support of this project, and especially production support. Andreas Kornmaaler Hansen has been an invaluable member of the research team and was a production assistant and videographer for the performance. Cassandra Sandu and Christina Duna facilitated the artistic workshops and enriched the academic discussions on the project. Special thanks to Liucija Paniuskyte for film editing. The initial drawing software was developed by David Hinwood and James Ireland from the University of Canberra in collaboration with Andreas Kornmaaler Hansen, Kristoffer Wulf Kristensen, and Jonas Ebler Pedersen. Finally, a special thanks to the Art and Technology students who shared their lab spaces with the research team. We also thank the Municipality of Aalborg and the Kulturvitaminer program for their participation in workshops.

REFERENCES

Bentley, T., Johnston, L., and von Baggo, K. (2005). "Evaluation Using Cued-Recall Debrief to Elicit Information about a User's Affective Experiences," in OZCHI '05 Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future, Canberra Australia, November 21 – 25, 2005 (Narrabundah, AUS: Computer-Human Interaction Special Interest Group (CHISIG) of Australia), 1–10.

Berio, D., Calinon, S., and Leymarie, F. F. (2016). "Learning Dynamic Graffiti Strokes with a Compliant Robot," in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), Daejeon, Korea (South), 9–14 October 2016. doi:10.1109/IROS.2016.7759586

Blackwell, T., Bown, O., and Young, M. (2012). "Live Algorithms: Towards Autonomous Computer Improvisers," in *Computers and Creativity*, Editors J. McCormack and M. d'Inverno (Berlin, Heidelberg: Springer), 147–174. doi:10.1007/978-3-642-31727-9_6

Boden, M. (1994). "Creativity and Computers," in *Artificial Intelligence and Creativity. Studies in Cognitive Systems,*. Editor D. Terry (Dordrecht: Springer) 17, 3–26. doi:10.1007/978-94-017-0793-0_1

Bown, O. (2012). "Generative and Adaptive Creativity: A Unified Approach to Creativity in Nature, Humans and Machines," in *Computers and Creativity*. Editors J. McCormack and M. d'Inverno (Berlin, Heidelberg: Springer), 361–381. doi:10.1007/978-3-642-31727-9_14

Casper, M. J., and Latour, B. (2000). Pandora's Hope: Essays on the Reality of Science Studies. *Contemp. Sociol.* 29, 754. doi:10.2307/2655272

Chung, S. (2018). *Issues in Science and Technology 35*. Arizona: Arizona State University 1. 2018 Data From.

Csikszentmihalyi, M., and Getzels, J. W. (2014). "Discovery-oriented Behavior and the Originality of Creative Products: A Study with Artists," in *The Systems Model of Creativity: The Collected Works of Mihaly Csikszentmihalyi* (Dordrecht: Springer Netherlands), 1–10. doi:10.1007/978-94-017-9085-7_1

Csikszentmihalyi, M. (1998). "Implications of a Systems Perspective for the Study of Creativity," in *Handbook of Creativity*. Editor R. J. Sternberg (Cambridge: Cambridge University Press), 313–336. doi:10.1017/CBO9780511807916.018

Cuan, C. (2021). Output: Choreographed and Reconfigured Human and Industrial Robot Bodies across Artistic Modalities. *Front. Robot. AI.* 7, 218. doi:10.3389/frobt.2020.576790

Dahlstedt, P. (2012). "Between Material and Ideas: A Process-Based Spatial Model of Artistic Creativity," in *Computers and Creativity*. Editors J. McCormack and M. d'Inverno (Berlin, Heidelberg: Springer), 205–233. doi:10.1007/978-3-642-31727-9_8

Demers, L.-P. (2016). *The Multiple Bodies of a Machine Performer*. Singapore: Springer Singapore, 273–306. doi:10.1007/978-981-10-0321-9_14

Derivative (2021). Data From: Touchdesigner.

Donnarumma, M. (2020). Across Bodily and Disciplinary Borders. *Perform. Res.* 25, 36–44. doi:10.1080/13528165.2020.1842028

Fdili Alaoui, S., Schiphorst, T., Cuykendall, S., Carlson, K., Studd, K., and Bradley, K. (2015). "Strategies for Embodied Design," in Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition, Glasgow, United Kingdom, June, 2015 (New York, NY, USA: Association for Computing Machinery), 121–130. doi:10.1145/2757226.2757238

Gannon, M. (2018). *Human-Centered Interfaces For Autonomous Fabrication Machines*. Pittsburgh, PA: Carnegie Mellon University.

Gemeinboeck, P. (2021). The Aesthetics of Encounter: A Relational-Performative Design Approach to Human-Robot Interaction. *Front. Robot. AI.* 7, 217. doi:10.3389/frobt.2020.577900

Giaccardi, E., and Karana, E. (2015). "Foundations of Materials Experience," in CHI '15Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul Republic of Korea, April 18 – 23, 2015 (New York, NY, USA. Association for Computing Machinery), 2447–2456. doi:10.1145/2702123.2702337

Gülzow, J., Grayver, L., and Deussen, O. (2018). Self-improving Robotic Brushstroke Replication. *Arts* 7, 84. doi:10.3390/arts7040084

Hayles, N. (2012). *How We Think: Digital Media and Contemporary Technogenesis*. London: University of Chicago Press.

Herath, D., Kroos, C., and Stelarc, C. (2016). *Robots and Art:Exploring and unlikely symbioses*. Singapore: Springer. doi:10.1007/978-981-10-0321-9

Hinwood, D., Ireland, J., Jochum, E. A., and Herath, D. (2018). "A Proposed Wizard of Oz Architecture for a Human-Robot Collaborative Drawing Task," in *Social Robotics*. Editors S. Ge, Shuzhi, C. Cabibihan, S. John-John, A. Miguel, J. Broadbent, H. Elizabeth, et al. (Cham: Springer International Publishing), 35–44. doi:10.1007/978-3-030-05204-1_4

If/then (2020). Data From: If/then.

Ingold, T. (2009). The Textility of Making. *Cambridge J. Econ.* 34, 91–102. doi:10.1093/cje/bep042

Jochum, E., and Derks, J. (2019). "Tonight We Improvise!," in MOCO '19: Proceedings of the 6th International Conference on Movement and Computing, Tempe, AZ, October, 2019 (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3347122.3347129

Jochum, E., Hansen, A. K., Duna, C., and Sandu, C. (2020). "Towards Creative Applications for Socially Assistive Robots. 3," in HRI'20; ACM/IEEE International Conference on Human-Robot Interaction, 23-03-2020 Through 26-03-2020.

Jochum, E., Millar, P., and Nuñez, D. (2017). Sequence and Chance: Design and Control Methods for Entertainment Robots. *Robotics Autonomous Syst.* 87, 372–380. doi:10.1016/j.robot.2016.08.019

Klee, P. (2013). Creative Confession. *Artist's Writings*. London: Tate Enterprises Limited.

Kluszczynski, R. (2010). Strategies of Interactive Art. *J. Aesthetics Cult.* 2, 5525. doi:10.3402/jac.v2i0.5525

Ladenheim, K., McNish, R., Rizvi, W., and LaViers, A. (2020). "Live Dance Performance Investigating the Feminine Cyborg Metaphor with a Motion-Activated Wearable Robot," in HRI '20: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot

Interaction, Cambridge, United Kingdom, March, 2020 (New York, NY, USA: Association for Computing Machinery), 243–251. doi:10.1145/3319502.3374837

Laviers, A., and Egerstedt, M. (2014). *Controls and Art: Inquiries and the Intersection of the Subjective and the Objective*. Switzerland: Springer. doi:10.1007/978-3-319-03904-6

Mamykina, L., Candy, L., and Edmonds, E. (2002). Collaborative Creativity. *Commun. ACM.* 45, 96–99. doi:10.1145/570907.570940

McCormack, J., and d'Inverno, M. (2012). *Computers and Creativity*. Berlin: Springer. doi:10.1007/978-3-642-31727-9

Nake, F. (2012). "Construction and Intuition: Creativity in Early Computer Art," in *Computers and Creativity*. Editors J. McCormack and M. d'Inverno (Berlin, Heidelberg: Springer), 61–94. doi:10.1007/978-3-642-31727-9_3

Pakrasi, I., Chakraborty, N., and LaViers, A. (2018). "A Design Methodology for Abstracting Character Archetypes onto Robotic Systems," in MOCO '18: Proceedings of the 5th International Conference on Movement and Computing, Genoa, Italy, June, 2018 (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3212721.3212809

Pedersen, J. E., Christensen, K. W., Herath, D., and Jochum, E. (2020). "I like the Way You Move: A Mixed-Methods Approach for Studying the Effects of Robot Motion on Collaborative Human Robot Interaction," in *Social Robotics*. Editors R. Wagner, R. Alan, Feil-Seifer, David, Haring, S. Kerstin, et al. (Cham: Springer International Publishing), 73–84. doi:10.1007/978-3-030-62056-1_7

Quantrill, M. (2002). "Integrating Computers as Explorers in Art Practice," in *Computers and Creativity*. Editors L. Candy and E. Edmonds (London: Springer London), 225–230. doi:10.1007/978-1-4471-0197-0_26

Riskin, J., and Bregović, M. (2017). *The Restless Clock: A History of the Centuries-Long Argument over what Makes Living Things Tick*. London: The University of Chicago Press, Ltd.

Salter, C. (2010). *Entangled: Technology and the Transformation of Performance*. Cambridge: The MIT Press.

Santos, M., Notomista, G., Mayya, S., and Egerstedt, M. (2020). Interactive Multi-Robot Painting through Colored Motion Trails. *Front. Robot. AI.* 7, 580415. doi:10.3389/frobt.2020.580415

Schacher, J., and Wei, L. (2019). "Gesture-Ink-Sound," in 6th International Conference on Movement and Computing (MOCO '19), Tempe, AZ, October, 2019 (New York, NY, USA: Association for Computing Machinery), 1–8. doi:10.1145/3347122.3347136

Schaffer, S. (1999). "Enlightened Automata," in *The Sciences in Enlightened Europe*. Editors W. Clark, J. Golinski, and S. Schaffer (London: The University of Chicago Press, Ltd), 136–165.

Sheets-Johnstone, M. (2011). Embodied Minds or Mindful Bodies? a Question of Fundamental, Inherently Inter-related Aspects of Animation. *Subjectivity* 4, 451–466. doi:10.1057/sub.2011.21

Shildrick, M. (2013). Re-imagining Embodiment: Prostheses, Supplements and Boundaries. *Somatechnics* 3, 270–286. doi:10.3366/soma.2013.0098

Simondon, G. (2016). *On the Mode of Existence of Technical Objects*. Minneapolis: Univocal Publishing.

Smith, G., and Leymarie, F. F. (2017). The Machine as Artist: An Introduction. *Arts* 6, 5. doi:10.3390/arts6020005

Spiel, K. (2021). "The Bodies of TEI - Investigating Norms and Assumptions in the Design of Embodied Interaction," in TEI '21: Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction, Salzburg, Austria, February, 2021 (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3430524.3440651

Stiegler, B. (1998). *Technics and Time 1: The Fault of Epimetheus, Trans*. Stanford: Stanford University Press.

Still, A., and d'Inverno, M. (2019). Can Machines Be Artists? a Deweyan Response in Theory and Practice. *Arts* 8, 36. doi:10.3390/arts8010036

Tafler, D. I. (2019). Drawing Spirits in the Sand: Performative Storytelling in the Digital Age. *Religions* 10, 492. doi:10.3390/rel10090492

Tresset, P., and Deussen, O. (2014). "Artistically Skilled Embodied Agents," in AISB 2014, London, April 1 – 4 2014 (UK: Goldsmiths, University of London).

Vorn, B. (2016). "I Want to Believe-Empathy and Catharsis in Robotic Art," in *Robots and Art: Exploring an Unlikely Symbiosis*. Editors D. Herath and C. Kroos Stelarc (Singapore: Springer Singapore), 365–377. doi:10.1007/978-981-10-0321-9_18

Wainer, J., Feil-Seifer, D., Shell, D., and Mataric, M. (2006). "The Role of Physical Embodiment in Human-Robot Interaction," in ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication, Hatfield, UK, October, 2006, 117122. doi:10.1109/ROMAN.2006.314404

Walter, B. (1996). "Paintings, or Signs and marks," in *Selected Writings, 1, 1913-1926*. Editors M. Bullock, W. Jennigs, and Michael (Cambridge: Harvard University Press), 83–86.

Wells, F. C. (2013). *The Heart of Leonardo*. London: Springer. doi:10.1007/978-1-4471-4531-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 02 August 2021
doi: 10.3389/frobt.2021.680586

Creativity in Generative Musical Networks: Evidence From Two Case Studies

Rodrigo F. Cádiz[1,2], Agustín Macaya[1], Manuel Cartagena[3] and Denis Parra[3]*

[1]Department of Electrical Engineering, Faculty of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile, [2]Music Institute, Faculty of Arts, Pontificia Universidad Católica de Chile, Santiago, Chile, [3]Department of Computer Science, Faculty of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile

Deep learning, one of the fastest-growing branches of artificial intelligence, has become one of the most relevant research and development areas of the last years, especially since 2012, when a neural network surpassed the most advanced image classification techniques of the time. This spectacular development has not been alien to the world of the arts, as recent advances in generative networks have made possible the artificial creation of high-quality content such as images, movies or music. We believe that these novel generative models propose a great challenge to our current understanding of computational creativity. If a robot can now create music that an expert cannot distinguish from music composed by a human, or create novel musical entities that were not known at training time, or exhibit conceptual leaps, does it mean that the machine is then creative? We believe that the emergence of these generative models clearly signals that much more research needs to be done in this area. We would like to contribute to this debate with two case studies of our own: TimbreNet, a variational auto-encoder network trained to generate audio-based musical chords, and StyleGAN Pianorolls, a generative adversarial network capable of creating short musical excerpts, despite the fact that it was trained with images and not musical data. We discuss and assess these generative models in terms of their creativity and we show that they are in practice capable of learning musical concepts that are not obvious based on the training data, and we hypothesize that these deep models, based on our current understanding of creativity in robots and machines, can be considered, in fact, creative.

Keywords: generative models, music, deep learning - artificial neural network (DL-ANN), VAE (variational AutoEncoder), GAN (generative adversarial network), creativity

Edited by:
Patricia Alves-Oliveira, University of Washington, United States

Reviewed by:
Lina Yao, University of New South Wales, Australia
Fabrizio Augusto Poltronieri, De Montfort University, United Kingdom

****Correspondence:***
Rodrigo F. Cádiz
rcadiz@uc.cl

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *15 March 2021*
Accepted: *08 July 2021*
Published: *02 August 2021*

Citation:
Cádiz RF, Macaya A, Cartagena M and Parra D (2021) Creativity in Generative Musical Networks: Evidence From Two Case Studies. Front. Robot. AI 8:680586. doi: 10.3389/frobt.2021.680586

1 INTRODUCTION

The field of deep learning (DL), one of the branches of artificial intelligence (AI), has become one of the most relevant and fast-growing research and development areas of recent times, especially since 2012, when an artificial neural network (ANN) called AlexNet (Krizhevsky et al., 2012) surpassed the most advanced image classification techniques to the date (Briot et al., 2020). This AI boom has happened because of three factors: first, today there is much more data available, second, there are much faster and more powerful computers available to researchers and third, technical advances. In particular, breakthroughs in the theory of ANNs, such as new training methods, convolutional networks, recurrent networks with short and long term memory, regularization techniques such as dropout, generative and transformer models, among others. These advances have allowed for the design and implementation of very sophisticated and complex AI models.

Indeed, DL models have been proven useful even in very difficult computational tasks, such as solving very difficult inverse problems with great precision (Goodfellow et al., 2016, 12). These approaches have the advantage that all parameters are objectively computed at the training stage, minimizing the error between predictions and the results provided by the training data. Training processes tend to be of high computational load, but once the training is finished, ANN-based reconstructions are extremely fast. However, classification and regression are perhaps not the most impressive applications of DL. There is increasing evidence showing that DL models can also generate very realistic audiovisual content, apparently at the same level of expert humans. In particular, variational auto-encoders (VAEs) and generative adversarial networks (GANs) are the most widely used generative strategies, yielding very interesting results, especially in the form of deep-fakes or deep video portraits (Kim et al., 2018).

Research in the field of robot musicianship has a rich history (Rowe, 2004) and it has experienced an increasing interest in recent times (Bretan and Weinberg, 2016). Currently there are robotic performers that can achieve very expressive performance levels, particularly with reinforcement learning approaches (Hantrakul et al., 2018) and machines that can compose music in real-time based on inference rules (Cádiz, 2020), or with direct interaction with its environment and people (Miranda and Tikhanoff, 2005). However, the question of creativity in robot musicianship remains elusive. We would like to contribute to the creation of better robotics composers or improvisers by studying the creativity of DL generative musical networks and identifying musical elements that could enlighten the discussion.

In this article, we study the use of generative models for musical content creation by means of a literature survey as well as by presenting two case studies and examining them under the light of computational creativity theory. The first use case describes the implementation and usage of a VAE model to encode and generate piano chords directly in audio, which we call TimbreNet. The second use case is a generator of musical piano rolls based on the StyleGAN 2 network architecture. Piano rolls are a widely used two-dimensional representation of musical data, very similar to a musical score in the sense that the *x*-axis represents time while pitches are encoded in the *y*-axis. We believe that both generative models, even though they have different architectures and music representations, exhibit behavior that could be classified as creative, as they can represent musical concepts that are not obvious based on the training data, and also exhibit conceptual leaps.

This article is structured as follows. In **section 2**, we discuss the most important generative models and show how they are able to create content. In **section 3**, we introduce the concept of computational creativity and provide a state-of-the-art review on the topic, including the most used ways for assessing creativity in computational systems. In **section 4**, we provide two case studies of generative networks that we think exhibit creative behavior. In **section 5**, we describe a simple perceptual survey we created to subjectively assess traits of creativity of the results of one of our models. In **section 6**, we discuss these case studies under the light of computational creativity theory and assess their creativity. Finally, in **section 7**, we present our main findings and layout ideas for future work.

2 DEEP GENERATIVE MODELS

According to Goodfellow et al. (2014), DL promises that we can build models that represent rich and hierarchical probability data distributions, such as natural images or audio, with great accuracy. This potential of DL makes perfect sense for music, being in essence very rich, structured, and also hierarchical information encoded in either a two-dimensional format (a score or a piano roll) or as one-dimensional array of audio samples. It is no surprise then that this amazing growth of DL in recent years has also greatly impacted the world of music and of machine musicianship.

As we stated before, perhaps one of the most interesting aspects that these networks can do now, apart from classification and regression, is the generation of content. In particular, ingenious network architectures have been designed for the generation of images, text, paintings, drawings or music (Briot et al., 2020). In the music realm, perhaps one of the most relevant research devoted to music generation is being carried out by the Magenta project.[1], a part of Google Brain. The goals of Magenta is not only to automatically generate new content, but to explore the role of ML as a tool in the artistic and creative process.

One of the most important aspects of generative DL approaches for music is their generality. As Briot et al. (2020) emphasize: "As opposed to handcrafted models, such as grammar-based or rule-based music generation systems, a machine learning-based generation system can be agnostic, as it learns a model from an arbitrary corpus of music. As a result, the same system may be used for various musical genres. Therefore, as more large-scale musical datasets are made available, a machine learning-based generation system will be able to automatically learn a musical style from a corpus and to generate new musical content". Contrary to rule-based structured representations, DL is very appropriate for handling raw unstructured data, and to extract higher-level information from it. We believe that this particular capacity makes DL a suitable technique for novel musical content generation.

Almost exclusively, these efforts aimed towards musical content creation are based on generative models, which are unsupervised models that intend to represent probability distributions over multiple variables (Goodfellow et al., 2016, 645). Some approaches estimate a probability distribution function explicitly, while others support operations that require some knowledge of it, such as drawing samples from the distributions. Although several models can generate content,

[1]http://magenta.tensorflow.org.

there are two that are the most promising and relevant today: VAEs and GANs (Charniak, 2018, 137).

VAEs are a probabilistic type of ANNs known as auto-encoders, which are functions whose output is nearly identical to the input (Charniak, 2018, 137). They are encoders because to generate the output, the network must have learned to represent the input data in a much more compact way, more specifically a low-dimensional space, known as a latent space. In a VAE, samples are drawn from the latent space to generate new outputs. As it is not possible to fill an entire latent space with only training data, some points in this space will inevitably generate outputs that were previously unknown to the network, an apparent sign of creativity.

More specifically, the loss function of a VAE (Kingma and Welling, 2014) can be described by the equation:

$$\mathcal{L}_{\mathrm{VAE}} = \mathbb{E}_{q(\boldsymbol{z}|\boldsymbol{x})}\left[\log p(\boldsymbol{x}|\boldsymbol{z})\right] - \mathrm{KL}\left(q(\boldsymbol{z}|\boldsymbol{x}) \parallel p(\boldsymbol{z})\right) \tag{1}$$

The first term in **Eq. 1** corresponds to the reconstruction loss, which is the expectation over the log-likelihood of the reconstructed data points using the decoder $p(x|z)$, where z is sampled from the encoder $q(z|x)$. The second part of the equation is considered a regularization term, the Kullback-Leibler (KL) divergence between the encoder distribution $q(z|x)$ and $p(z)$. The prior distribution is placed over the encoder and decoder parameters and in this article we use a Gaussian prior with mean zero and variance one, since it facilitates the generation of new samples from the latent space, it has an analytical evaluation of the KL divergence in the loss function, and the non-linear decoder can mimic arbitrarily complicated distributions if necessary starting from the prior Gaussian distribution. This loss function of VAEs decreases as the input and output data are alike, and in every iteration, a VAE network learns to represent the input space in a more efficient and compressed form. The decoder part of the network can thus generate novel output that share a lot of the characteristics of the input space.

Another promising research in DL is related to the development of Generative Adversarial Networks (GANs), which represent a significant shift from traditional DL architectures. In GANs, two ANN work against each other in adversarial training to produce generative models (Kalin, 2018, 9). More formally, GANs (Goodfellow et al., 2014) have provided a new framework for estimating generative models via what is called an adversarial process, in which two models are simultaneously trained. In this approach, the input data distribution is estimated by a generative model (G), while a discriminator model (D) evaluates the probability that a freshly generated output provenances is indeed from the training data rather than from the generator G. The whole idea of this approach is to make the generative model G so good that eventually D might be fooled by a false input. If this happens, it means that G is generating fake data that is indistinguishable from real data, also a possible indication of creativity.

This process can be summarized in **Eq. 2**, where the goal with the adversarial training is to find the function D which maximizes the log probability of correct cases, while the generator G minimizes the log-probability of the discriminator being correct.

$$\min_{G}\max_{D} V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}\left[\log D(\boldsymbol{x})\right] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[\log\left(1 - D(G(\boldsymbol{z}))\right)\right] \tag{2}$$

Once trained, these networks can convert random noise into highly realistic content, such as images or audio signals. There are several advantages of this approach: GANs generalize well with limited data and they can conceive new scenes from small datasets, but perhaps, the most important aspect is that they make simulated data look highly realistic (Kalin, 2018, 10).

2.1 Musical Generative Networks

In the musical field, generative models such as the ones we previously discussed have been gaining popularity in recent times for the creation of audible content. We now provide a literature review of the most relevant works for music creation based on these two architectures.

Hadjeres et al. (2016) created DeepBach, a neural network capable of modeling polyphonic music and pieces in the anthem genre, which harmonizes Bach-style choral in a very convincing way. Oord et al. (2016) created Wavenet, a network that renders audio files at the sample level. Wavenet has been shown to produce good results in human voice and speech. Engel et al. (2017), using NSynth, a very large dataset of sound for digital synthesis, were able to improve both the qualitative and quantitative performance of WaveNet. Their model learns a manifold of embeddings that allows for instrument morphing, a meaningful way for interpolating timbre that results in new types of realistic and expressive sounds. Sturm et al. (2016) have used generative models for music transcription problems. They specifically designed generative long short-term memory (LSTM) models, for the task of music transcription and composition. Roberts et al. (2018) created MusicVAE, a network designed for the generation of compact latent spaces that can be later interpolated for the generation of content. Yang et al. (2017) created MidiNet, a convolutional adversary generation network able to produce melodies in the MIDI format. Dong et al. (2018a) created MuseGAN, an adversarial network for symbolic music and accompaniment, in this case in the rock genre. Roberts et al. (2017) designed a VAE for the generation of a variety of musical sequences at various bar scales: 2-bar, 16-bar or 32-bars. Yamshchikov and Tikhonov (2020) propose a novel DL architecture labeled as Variational Recurrent Autoencoder (VRASH), that used previous outputs as additional inputs, forming a history of the analyzed events. VRASH "listens" to the notes already output and uses them as a feed for "historic" input. This is the first application of such a generative approach to the generation of music rather than text. Weber et al. (2019) were able to generate novel melodies via a ANN model that ensures, with high probability, consistency of melody and rhythm with a target set of sample songs. A unique aspect of this work is that they propose the usage of Perlin noise in opposition to the more widely used white noise in the context of VAEs.

In the field of audio processing, impressive advances have been made in the last two years. As an example, we can cite Spleeter (Hennequin et al., 2019), a music source separation tool for up to five simultaneous voices based on deep learning. This task is extremely hard when tackled with traditional signal processing approaches. Another interesting example is the Differentiable Digital Signal Processing (DDSP) library (Engel et al., 2020), created by Magenta, which enables direct integration of classic signal processing elements with the power of deep learning. This approach achieves high-fidelity audio generation without the need for large models or adversarial architectures. DDSP models are similar to vocoder systems, which are physically and perceptually motivated, and directly generate audio with oscillators, and do not work by predicting waveforms or Fourier coefficients, as traditional methods do.

In **section 4** we will elaborate on two generative models that we have built aimed towards the generation of audio-based musical chords and symbolic piano roll-based short musical sequences. In the specific case of chords generation, a significant amount of research is aimed towards chord recognition (Humphrey et al., 2012; Zhou and Lerch, 2015; Deng and Kwok, 2016; Korzeniowski and Widmer, 2016), in detriment of chord generation. The first study case what we present below is based on GanSynth (Engel et al., 2017), a GAN model that when its latent vector is sampled, it generates a complete audio excerpt, allowing for a smooth control of features such as pitch and timbre. Our model is based on GanSynth, but it was tuned for the specific case of chord sequences. In terms of piano roll sequence generation, MuseGAN (Dong et al., 2018a) is probably the most well-know model targeted for this specific musical format. Our second case study uses piano rolls instead of images in a network previously trained with only real-world images.

3 COMPUTATIONAL CREATIVITY

A very important question in the field of artificial intelligence is whether computers or robots can be creative. This is a very difficult research topic, as scientists have only embraced the study of human creativity in recent times (Sawyer, 2006, 3). According to Brown (1989), four distinct approaches have dominated the study of creativity: 1) an aspect of intelligence; 2) a largely unconscious process; 3) an aspect of problem-solving; and 4) an associative process. Nowadays, the study of creativity in humans has settled into what is called the socio-cultural approach, an interdisciplinary effort to explain how people are creative and their social and cultural contexts (Sawyer, 2006, 4).

It is a consensus that creativity can be defined as “the ability to generate novel, and valuable, ideas” (Boden, 2009). This definition implies the generation of “something that is both original and worthwhile” (Sternberg and Sternberg, 2012), or a “conceptual leap” by the combination of existing knowledge (Guzdial and Riedl, 2019). These “ideas” or “somethings” can take the form of intangibles, such as a scientific theory, a mathematical theorem, a musical composition, a neural network, a poem, or a joke; or even tangible physical objects, such as an invention, a robot, a mechanical tool, a chemical, a printed literary work, a sculpture, a digital circuit, or a painting. The notion of novelty is crucial for this understanding of creativity. But in addition, as previously stated, Boden (2009) emphasizes that creativity should be “valuable”. This implies a subject-dependent evaluation, as what influences the assessment we make of something is not only its features or objective properties, but rather how such a thing is produced and presented (Moruzzi, 2018). It is also worth emphasizing that novelty often implies unpredictability and uncertainty, especially in the case of musical creativity (Daikoku et al., 2021).

Carnovalini and Rodà (2020) observe that “the usual experience with machines is that we humans give a set of instructions to the machine along with some initial data (the input), and we expect the machine to behave in a way that is fully deterministic, always giving the same output when the same input is given”. This idea of deterministic robots is apparently very opposed to the whole notion of creativity, which supposes something novel and valuable. This notion of “novelty” is understood by Grace and Maher (2019) as “violated-expectations” models. However, as Mumford and Ventura (2015) point out, a “common misconception among non-specialists is that a computer program can only perform tasks which the programmer knows how to perform (albeit much faster). This leads to a belief that if an artificial system exhibits creative behavior, it only does so because it is leveraging the programmer’s creativity”.

There are other ways of conceptualizing creativity. In particular, the categories of combinatorial, exploratory and transformational creativity, proposed by Boden (2004), are very enlightening. The first one is about making unfamiliar combinations of known ideas. The second one involves a structured conceptual space that is explored. The third category implies changing this conceptual space allowing new ideas to become possible. All of these categories are related to the conceptual leaps proposed by Sternberg and Sternberg (2012) in different degrees.

Another important aspect of creativity is the ability to autonomously evaluate outcomes, to “know when to stop” (Moruzzi, 2021). This aspect of creativity is crucial to determine whether the produced outputs work or not and reminds us that the process of creativity requires hard work, that it does not happen by pure magic. This autonomy means that the creative agent should be the one performing the assessment, without external influence.

Creativity is usually attributed to humans. However, as Park (2019) asks: When we regard something as artwork, should it be exclusively created, selected, and combined by human beings? We are used to the idea that humans can create things or ideas that other humans judge to be “new”–this happens almost every day in every domain. But computers can also produce outputs that can be thought of being new. For example, Cope (1996) developed computer algorithms which he labeled as “Experiments in Musical Intelligence (EMI)”, that allowed computers to generate novel compositions in a particular musical style, two decades before the rise of deep learning techniques. It is no surprise, then, that the study of the phenomenon of creativity has been extended to computers and machines, under the label "Computational creativity", which is a field of inquiry seeking the modeling, simulation, or replication of creativity inside a computer. This field is interdisciplinary by nature, with links to traditional fields such as artificial intelligence, psychology, the arts, or philosophy. It is also known as creative computation, creative computing, or artificial creativity.

The goals of computational creativity are not only to design and build computer systems capable of achieving or enhancing human-level creativity, but also to better understand how human creativity works. In the particular case of deep generative networks, one of the most interesting and current theoretical research trend is to determine if these generative networks are creative or not and to what extent. Karimi et al. (2018) define creative systems as those intelligent systems that are capable of performing creative tasks in isolation or collaboration with other systems. These systems are creative because their results are judged as such by their human counterparts (Colton et al., 2015; Elgammal et al., 2017). There even exist Turing-style tests to assess creativity from machines that create artworks, by asking machines to create art that is indistinguishable from human-created works.

The question of how can machines and robots be creative is far from settled. On the one hand, there are authors, such as Hertzmann (2018), who argue that the current AI technology is not yet able to create since to do it requires "intention, inspiration, and desire to express something". However, the advances in AI open for music, as did photography with paintings more than 100 years ago, the possibility of generating new forms of artistic creation. It is possible to understand AI as a technology that can increase and enhance human capabilities (Carter and Nielsen, 2017). On the other hand, authors such as Elgammal et al. (2017) have no problem in considering their systems creative. As evidence, they have created an architecture of ANNs labelled CAN (Creative Adversarial Networks), which can look at visual art and learn the artistic style inherent in the works with which they were trained. Then, by modifying certain parameters of the network, the authors argue that they become creative because they are capable of generating new art that deviates from the styles that were previously learned. Similarly, Guzdial and Riedl (2019) present a novel training method for neural networks called Combinets, a more general approach for reusing existing trained models to derive new ones without retraining via recombination. In a sense, they can make a DL network "creative", in the sense that it is able to represent new knowledge as a combination of particular knowledge from previous existing cases. Another important evidence towards the existence of creativity in machines is presented by Wyse (2019), who examined five distinct features typically associated with creativity, and provided examples of mechanisms from generative DL architectures that give rise to each of these characteristics, producing very strong evidence in favor of DL architectures being creative.

Another unresolved topic in the computational creativity field is the evaluation of generative systems in terms of their creativity (Ritchie, 2019). As Moruzzi (2018) illustrates: "The subjective judgments and biases which come with the evaluation of something as creative make it impossible to objectively answer the question "Can a computer be creative?" What we are measuring when we provide an answer to this question, in fact, are not the computer's accomplishments but instead our subjective evaluation of them. We can then try to analyze not just the creativity exhibited by the outcome produced by the computer but, instead, the intention of the computer in producing it. In other words, we can judge whether the computer produced its outcome intentionally, i.e., consciously intending to produce exactly that outcome. We should then rephrase the question and ask: "Can a computer be intentionally creative?"".

We have identified four strategies for the evaluation of creativity in robots and algorithms reported in the literature. The first one follows what Jordanous (2019) calls a "creative-practitioner-type approach, producing a system and then presenting it to others, whose critical reaction determines its worth as a creative entity", even in real-time (Collins, 2007). The second one is described by Carnovalini and Rodà (2020): "have the author of the system describe the way it works and how it can be considered creative or not, and to what degree." A third one is to evaluate artificially generated music in a concert setting, just as normal auditors would assess a live musical situation (Eigenfeldt et al., 2012; Sturm et al., 2018), or in a museum-like setting for the case of the visual arts (Edmonds et al., 2009). Finally, a fourth approach that we can identify is described in (Yang and Lerch, 2020), who propose "informed objective metrics" to complement a subjective evaluation by a human. For example, some metrics can determine how well a computer-generated music can "fit" a particular musical genre.

In the following section we will describe two case studies and discuss their creativity under the light of what we have presented in this section. In particular, we will focus on the aspects of novelty, in the sense that the models produce something that is not expected, value, by assessing whether novel outputs make sense and function well in their context, and conceptual leaps, understood as the reuse of a particular type of knowledge to produce a different kind. For the purposes of this article, we will be using the second evaluation strategy, as we are the authors of both models.

4 CASE STUDIES

We now present two case studies: TimbreNet, an ANN based on the architecture of GanSynth (Engel et al., 2017) that can generate novel chords directly in audio format and StyleGAN Pianorolls, a generative model, based on StyleGAN 2 (Karras et al., 2020b), that can create novel musical excerpts in the form of piano rolls.

4.1 TimbreNet: A Creative Chord Generator

The network architecture is presented in **Figure 1**. Our design goal was to generate a useful tool for musical composition, by means of the latent space exploration. A VAE-based model can accept inputs directly from the user in contrast to GAN-based

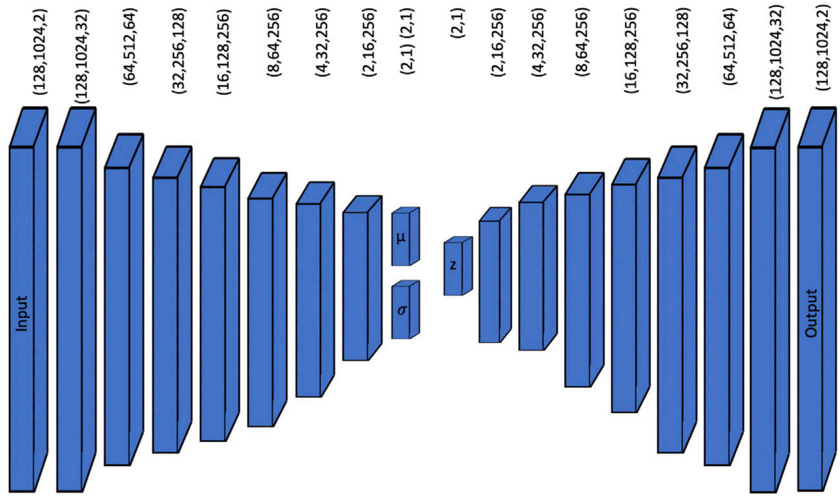


FIGURE 1 | Architecture of our VAE model for chord synthesis for the case $L = 2$. The encoder takes a (128,1024,2) MFCC image and passes it through several downsampling layers until it compacts the data into a low-dimension latent space z. The decoding process samples the latent vector using a Gaussian distribution of mean μ and standard deviation σ, and passes it through several upsampling layers until a (128,1024,2) output is obtained that is later converted to an audio signal.

models where the input is random noise. Although it is possible to mimic this behavior with conditional GANs, we opted for a VAE to obtain an explicit latent space representation of the input data. We based the encoder architecture on the discriminator structure of GanSynth (Engel et al., 2017) and the decoder architecture from its generator.

The encoder takes a MFCC (Mel Frequency Cepstral Coefficients) image of dimensions (128,1024,2) and passes it through one two-dimensional convolution layer with several additional filters generating a (128,1024,32) output that is fed to a series of 2 two-dimensional convolution layers with the same size padding and a Leaky ReLU non-linear activation function in cascade with 2×2 downsampling layers. This process keeps halving the images' size and duplicating the number of channels until a (2,16,256) layer is obtained. Then, a fully connected layer outputs a $(2L,1)$ vector, the latent space, that contains L means and L standard deviations for posterior sampling. We trained models with different sizes for L (specifically 3, 4, 8, 16, and 32), which is a meta-parameter that determines the dimension of the latent space. **Figure 1** displays the network structure for the case $L = 2$.

The sampling process begins with a $(L,1)$ mean vector and a (L,L) standard deviation diagonal matrix that is used for sampling the latent vector z from a normal distribution with mean μ and standard deviation σ. The z latent vector is fed to the decoder in cascade with a fully connected layer that generates a (2,16,256) output that then is followed by a series of two transposed convolutional layers in series with an 2×2 upsampling layer that keeps doubling the size of the image and halving the number of channels until a (128,1024,32) output is achieved. This output passes through a final convolutional layer that outputs the (128,1024,2) MFCC spectral representation of the generated audio. This spectral representation can be converted into an audio excerpt by inverse MFCC and STFT transformations.

4.1.1 Dataset and Model Training

Our dataset consisted on 43,200 recordings of tertian triads played at different keys, dynamic levels and octaves, performed by the second author on a piano. A triad is a chord containing three notes and a tertian chord is constructed by adding up notes separated by a major or minor third. Each recording was done in Ableton Live with a duration of 4 s, and a 16 kHz sampling rate. Piano keys were pressed for 3 s and then released during the last second. This dataset format has the same structure as the one used in Engel et al. (2017).

The base notes of the chords were the twelve notes of the western musical scale across three octaves giving a total of thirty-six base notes. For each base note, we recorded four different types of triads (major, minor, augmented, and diminished). We also recorded chords at three different levels of dynamics: f (forte), mf (mesoforte) and p (piano). For each combination, we produced ten different recordings for data augmentation purposes, as each recording is not an exact repetition of any other one, producing a total of 4,320 data examples and then we used data augmentation techniques to have a total of 43,200 examples. This dataset can be downloaded from the github repository of the project.[2].

We decided to use an MFCC representation of the audio samples for the input and output data, a design decision that has been proven to be very effective when working with convolutional networks designed for audio content generation (Engel et al., 2017). Magnitude and unwrapped phase appear codified in different channels of the image.

Figure 2 displays the MFCC transform of a 4-s audio recording of a piano chord performed *forte*. **Figure 3** displays the MFCC representation of a 4-s audio recording of the same

[2]https://github.com/CreativAI-UC/TimbreNet/tree/TimbreNet2/datasets/pianoTriadDataset.

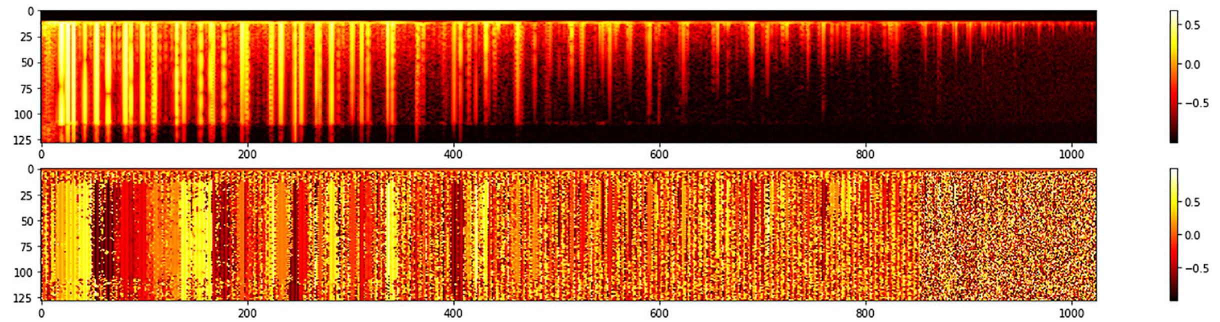


FIGURE 2 | MFCC representation of a *forte* chord used for training. The horizontal dimension represents time while the vertical dimension encodes frequency coefficients. Brighter yellow colors represent higher sound intensities. The top graph shows the magnitude of the frequency representation and the bottom displays its phase.

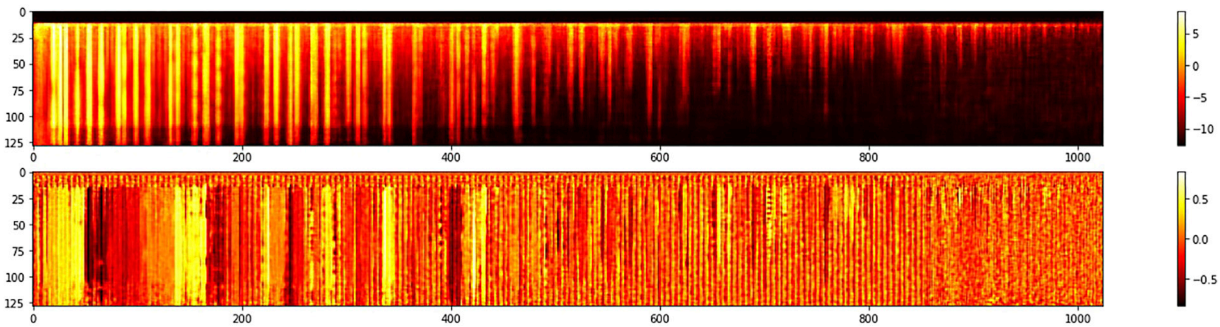


FIGURE 3 | MFCC representation of the same forte chord of **Figure 2** generated by the network's decoder. The horizontal dimension represents time while the vertical dimension encodes frequency coefficients. Brighter yellow colors represent higher sound intensities. The top graph shows the magnitude of the frequency representation and the bottom displays its phase.

forte chord of **Figure 2**, but in this case, the chord was generated by the network by sampling a trained position in the latent space, the one where the original chord can be found. In both figures, 2 and 3, magnitude is shown on the top half while unwrapped phase is displayed at the bottom part.

We used Tensorflow 2.0 to implement our model. For training, we split our dataset leaving 38,880 examples for training and validation, and 4,320 examples for testing. We used an Adam optimizer with default parameters and learning rate of 3×10^{-5}. We chose a batch size of 10, and the training was performed for a total of 250 epochs. The full training was done in about 6 h using one GPU, a Nvidia GTX 1080Ti. We used the standard cost function for VAE descried in **Eq. 1**, but in practice the model was trained to maximize the ELBO (Evidence Lower BOund) as proposed by Kingma and Welling (2014); Ranganath et al. (2014). We divided the 250 training epochs in five groups of 50 epochs. We started with a high reconstruction loss factor for the first 50 epochs and we decreased this factor across each epoch group. The high reconstruction loss factor allows for a good audio quality and then the later low reconstruction loss factor orders and clusters the latent space without a loss in audio quality (Higgins et al., 2017).

4.1.2 Latent Space

Figure 4 displays a three dimensional latent space generated by the network. On a macro level, chords are separated according to dynamic level as it can be observed on the right-most figure. On a micro level, chords are grouped with other chords with the same notes, and the nearest neighbors corresponds to the chords which have the most notes in common. This particular configuration of the latent space is very interesting from a musical stand point, as it appears that the networks learned to order the space based on musical concepts that are very fundamental such as common voicing, loudness and pitch.

One of the nice properties of latent spaces happens when one samples the space in an untrained position, a point in the plane that has not been previously trained by the network. In **Figure 5** we show the MFCC coefficients of a completely new chord generated by the network. Since different chords are clustered in the latent space, it is interesting to listen to chords that are generated in the space between clusters. We find out that the model is able to generate new chords with musical meaning that the model has never seen in the training dataset. **Figure 6** shows some examples of new chords generated by the

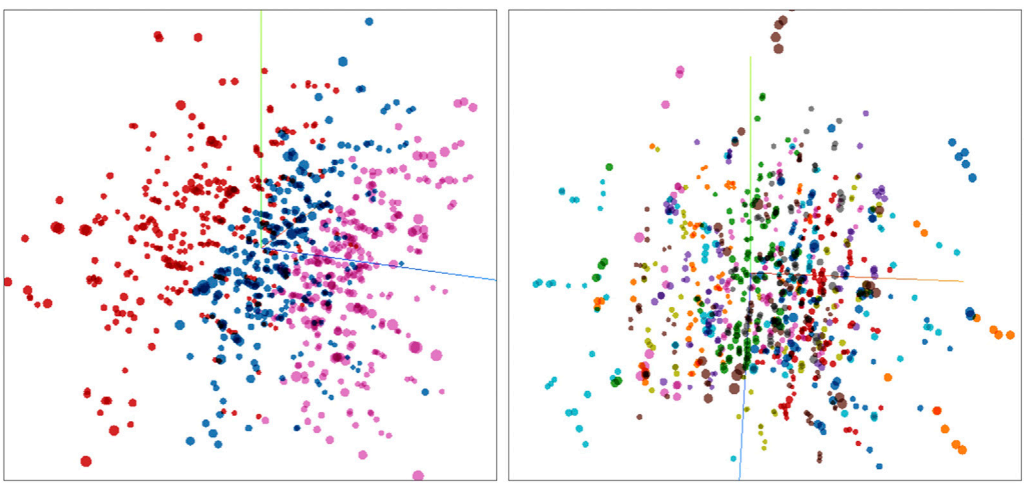


FIGURE 4 | Three dimensional latent space representation of the input dataset. Left: Chords colored by volume (forte, mesoforte, and piano). Right: chords colored by base notes. It can be appreciated that the latent space is segmented based on relevant musical properties.

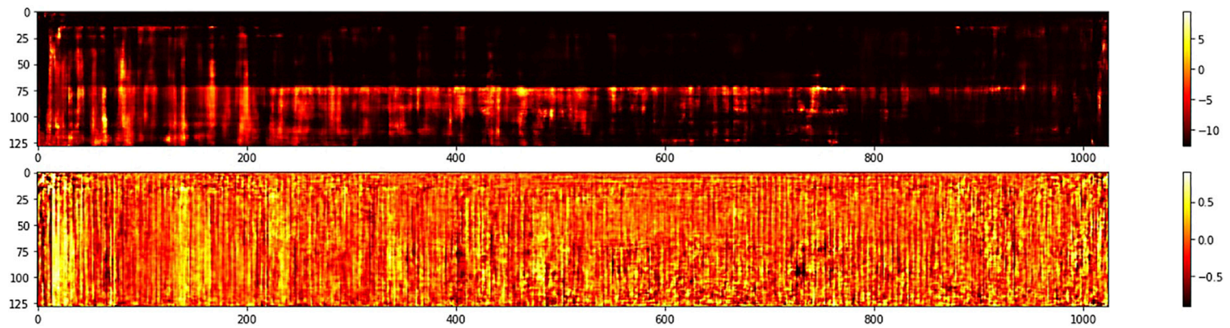


FIGURE 5 | MFCC of a new chord generated by the network's decoder by sampling an untrained point in the latent space. The horizontal dimension represents time while the vertical dimension encodes frequency coefficients. Brighter yellow colors represent higher sound intensities. The top graph shows the magnitude of the frequency representation and the bottom displays its phase.

network. The top three chords can be found in the dataset while the bottom three chords are four note chords that the model has never seen before during training.

We have created an interactive web-based tool for the exploration of this latent space, called Timbreplay.[3], in the same spirit of Moodplay (Andjelkovic et al., 2016, 2019) One nice feature of this tool is the generation of chord trajectories than the user can save for later use in musical compositions. In addition, audio examples of TimbreNet can be listened in the repository of the project.[4].

[3]http://timbreplay.ml.

[4]https://github.com/CreativAI-UC/TimbreNet/tree/TimbreNet2/generated_chords/paper_examples.

4.2 StyleGAN Pianorolls: A Creative Musical Excerpts Generator

Our second case study is based on the newly developed StyleGAN 2 (Karras et al., 2020b), which achieved state-of-the-art results in image generation, specifically on creating human faces that do not exist, but look highly realistic. Considering this results for generating images, which are a 2D representation of visual information, we experimented to see if this network could generate piano rolls, which are also a 2D representation, but in this case of a musical composition, with one axis representing time, and the other representing pitches, as it can be seen in **Figure 7** where both the input and output of the network are piano rolls represented as binary images. Even though we are fully aware that piano rolls are conceptually very different from human faces, we wanted to see if certain properties of visual information that they might have in common could be useful for training a musical generator model.

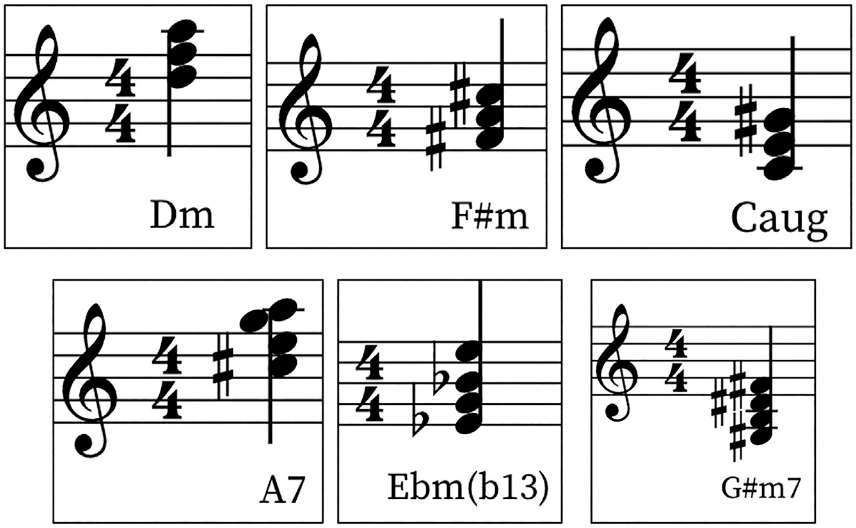


FIGURE 6 | New chords generated by the TimbreNet model. The top three chords are new but they are similar to chords that can be found in the training set. The bottom three chords are completely novel, with different number of notes and representing different tonal functions, such as a dominant seventh or minor-minor seventh chords..

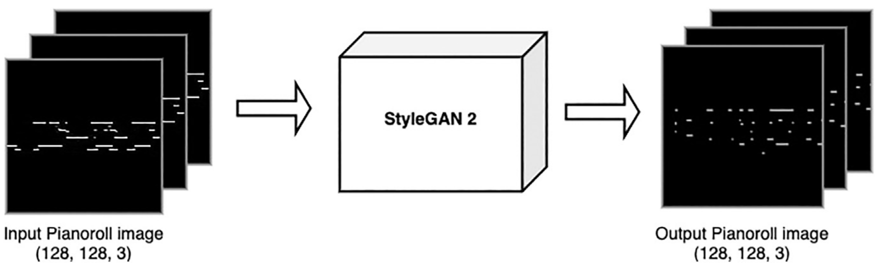


FIGURE 7 | Representation of input and output of the StyleGAN 2 network with piano rolls. A symbolic representation of a musical excerpt in the form of a 128 × 128 × 3 piano roll is used to train the network. The output is another piano roll, which can later be transformed into midi or audio.

4.2.1 Dataset and Model Training

We used the implementation of StyleGAN2-ADA in Tensorflow 1.14 provided by Karras et al. (2020a) that has an adaptive discriminator augmentation to better train with limited data. As inputs, we used the MAESTRO dataset V2.0.0 (Hawthorne et al., 2018), that consists of over 200 h of piano performances, which include raw audio and midi, 1,282 in total. We only used midi files, which can be very easily transformed to piano rolls. For each performance, its midi file was binarized and split into segments of 4 bars divided into 32 time steps each. After removing empty splits this processing resulted in 269,719 pianoroll images of shape (128, 128), this decision was made because the StyleGAN architecture has a constraint of using squared shape images. Although this constraint implies shorter musical segments, there's still interesting information to be captured in the training data. We used the same loss function of **Eq. 2**. The model was trained on a Tesla V100 in Google Colab from a previously trained checkpoint on the FFHQ Dataset (Karras et al., 2018) which consists of human faces from Flickr. Surprisingly, even though the network previously knew human faces only, it was relatively easy to have it recognize and generate musical excerpts, as we detail below.

4.2.2 Latent Space

One of the creative features of using the StyleGAN 2 architecture is that its random noise input is mapped into a disentangled latent space, called the w-space, through multiple fully connected layers. This new latent space is much richer to explore than the traditional latent space usually used in GANs, known as the z-space. The objective of using this disentangled w-space was to better separate different characteristics of the network's output, allowing a much finer control of the generation process when producing new content.

For notated music this space has a lot of potential for further exploration. For example, one appealing idea is finding trajectories in the latent w-space that can change a specific characteristic of the output without changing other features, which means keeping other musical features constant. Some

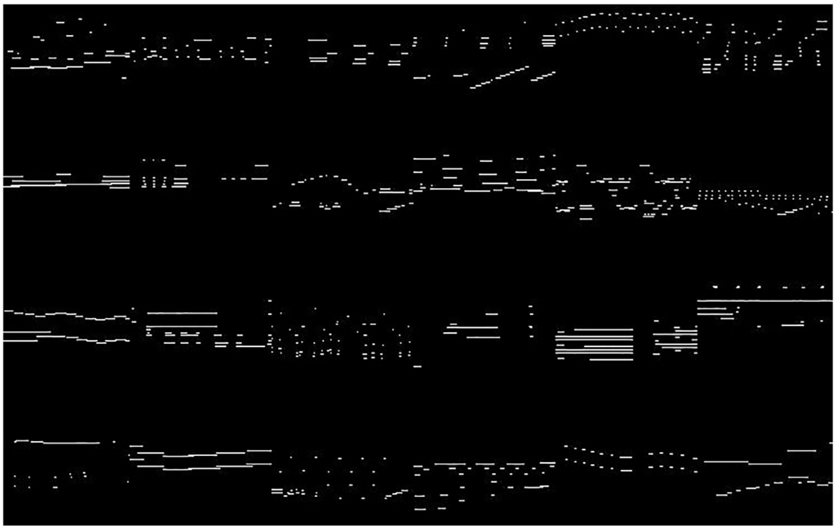


FIGURE 8 | 24 examples of real piano rolls used to train StyleGAN Pianorolls arranged in 4 rows and 6 columns. The examples exhibit great variation in their musical structure.

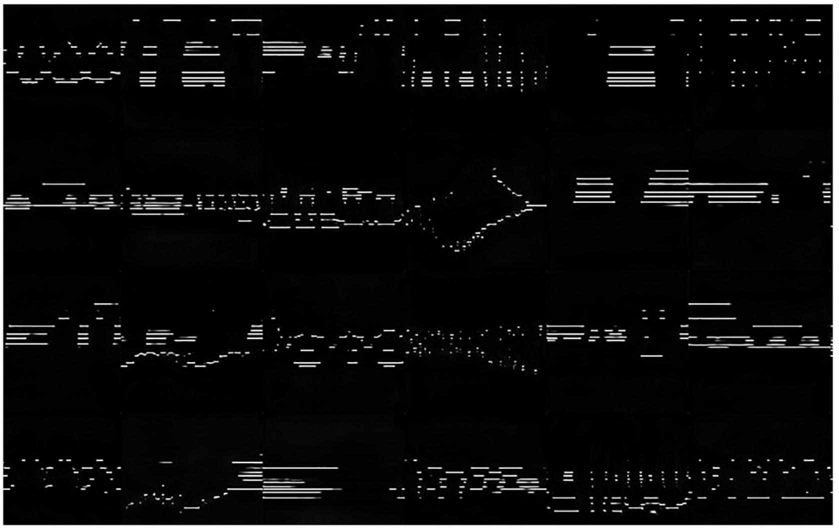


FIGURE 9 | 24 examples of fake piano rolls generated by StyleGAN Pianorolls arranged in 4 rows and 6 columns. The generated excerpts exhibit great variation in their musical structure, as it is the case of the input data.

examples of desired musical changes can be the number of pitches in the excerpt, its tonal key, the amount of silences, or the amount of polyphony, among other interesting musical features that can be described in a piano roll representation.

In **Figures 8**, **9** there's a comparison of several real input images against fake ones that were generated by our network. A first visual inspection of the images reveal that the fake images look very similar to the real ones. In terms of musical structures and motifs. the network is able to generate a great variety of musical ideas, ranging from pointillistic short events, as it can be observed in **Figure 10A**, to long chordal structures such as in **Figure 10F**. By interpolation of the latent space, it is also possible to generate a musical progression from one sample to another, with a variable number of intermediate steps, as **Figure 10** depicts. For a more musical evaluation, we published a folder.[5] with some selected samples to examine what this approach can potentially generate. There are single samples, which are the direct output of the network translated to MIDI and transformed

[5]https://drive.google.com/drive/folders/1cj-Y38GMxg4m0REWyU3ZvIUtWtp-TiOS?usp=sharing.

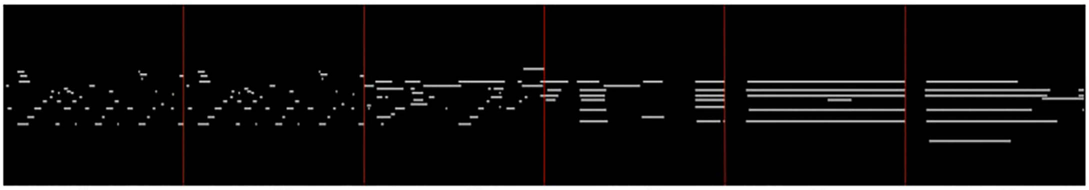


FIGURE 10 | StyleGAN Pianorolls is able to generate a variety of musical ideas **(A–F)**. The latent space can also be interpolated between 2 outputs to generate a musically-meaningful sequence. In this case, the generated sequence exhibits how the network morphs from sample A to sample F in 4 steps, visually divided by a red line for easier differentiation.

to an audio file using Timidity++, and also sequence samples that are the concatenation of multiple outputs interpolated from two points in the *w*-space, further explanation of the types of generation will be explained in the next section.

4.2.3 Generation

The process of generating an audio file from the output image of StyleGAN 2 has two parts: 1) defining a threshold and tempo for the generated piece, and 2) transforming the image to a numerical matrix. The first step is needed because the model returns a grayscale image, where the pixel values are between 0 and 255, and has three channels. With the threshold defined, we took the mean of the three channels and binarized the image using this threshold to determine which pixels correspond to played notes, thus, obtaining the numerical matrix which can be transformed to a pianoroll using the pypianoroll package developed by Dong et al. (2018b) to later convert it to a MIDI file. For listening to these files we used Timidity++ to convert them to a wave file.

We can generate new musical excerpts using this model through exploration of the latent *w*-space, changing the input values to get new pieces. Another interesting musical application is to interpolate between two examples generated by the network, defining the number of steps we can generate a sequence of concatenated outputs while moving from one point in the latent space to another, as shown in **Figure 10**. In the supplemented folder there's examples of different sequences from two random sampled points in the latent space, showing how the trained model evolves one excerpt into another in a series of steps.

5 PERCEPTUAL EVALUATION

We designed a very simple survey to obtain a first approximation to the perceptual validity of our results aimed towards determining whether our StyleGAN piano rolls network was able to generate musical excerpts that could be judged to be creative by human beings. Given the well-known ability of GANs to create realistic portraits, we created two videos.[6] based on StyleGAN2 content. The visual content was generated by a StyleGAN 2 neural network trained with publicly available images of portraits of the Chilean National Art Museum (Museo Nacional de Bellas Artes), in the same fashion described in Karras et al. (2020b). The audio content was generated using two versions of the StyleGAN piano rolls model, one trained with different instruments from the LAKH MIDI Dataset (Raffel, 2016), and the other with the MAESTRO dataset (Hawthorne et al., 2018). We curated different musical excerpts from these networks to assemble the complete musical pieces. Our work consisted mainly in organizing the different fragments generated to create longer structures with multiple instrumentations, instead of focusing on a single instrument. It is important to clarify that the audiovisual content was completely generated by StyleGAN 2 networks, that none of those faces that appear in video exist in reality and neither do the musical structures that can be heard, they were completely created by a machine. Only the temporal organization of the music was done with human intervention. Finally, to achieve the final audiovisual results we used the Lucid Sonic Dreams.[7] library, which uses a StyleGAN2 model to explore its latent space by synchronizing the transitions with a given audio, creating interesting movements to the rhythm of the music.

We asked participants to assess the creativity of each of the videos in terms of their audiovisual, visual only and audio only content, by selecting a number in a Likert scale from 1 to 5. 1 corresponded to the label "Disagree", while five indicated "Agree". 3 indicated no preference towards any side. For both videos, we evaluated the level of agreement/disagreement with the following statements:

1) The audiovisual content of the video is creative
2) The visual content of the video is creative
3) The audio content of the video is creative

Forty-four participants responded the survey over the internet. The results are shown in **Figures 11**, **12**, respectively. It is very clear, for both videos, that the majority of the subjects were in agreement with the statement that the content was creative. All three type of contents: audiovisual, visual only and audio only, especially in the second video, were judged to be creative by a great majority of participants.

[6]https://youtu.be/THxEOvRG4Ss, https://youtu.be/6CVAzQiMPlM.

[7]https://github.com/mikaelalafriz/lucid-sonic-dreams.

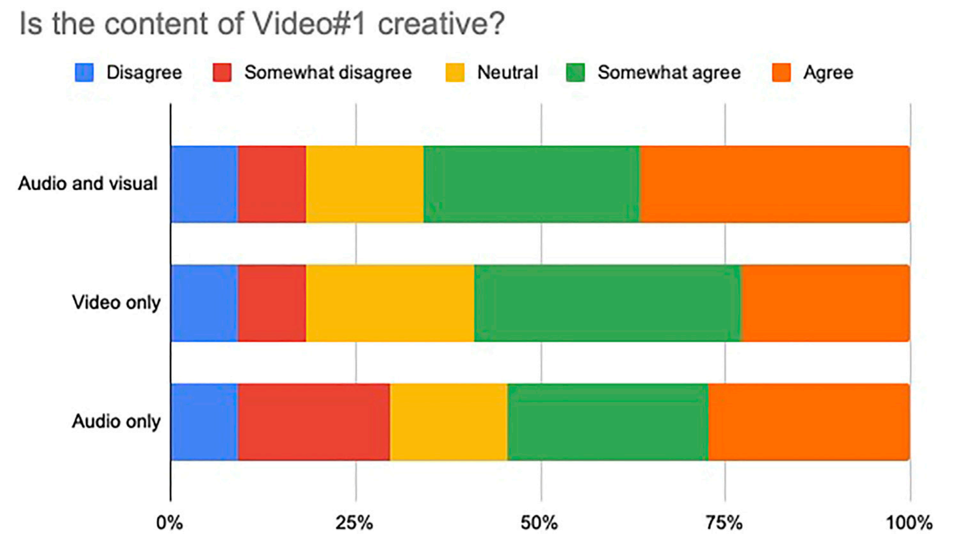


FIGURE 11 | Perceptual evaluation of Video #1. Forty-four participants responded the survey over the internet. All three type of contents: audiovisual, visual only and audio only, were judged to be creative by a great majority of participants.

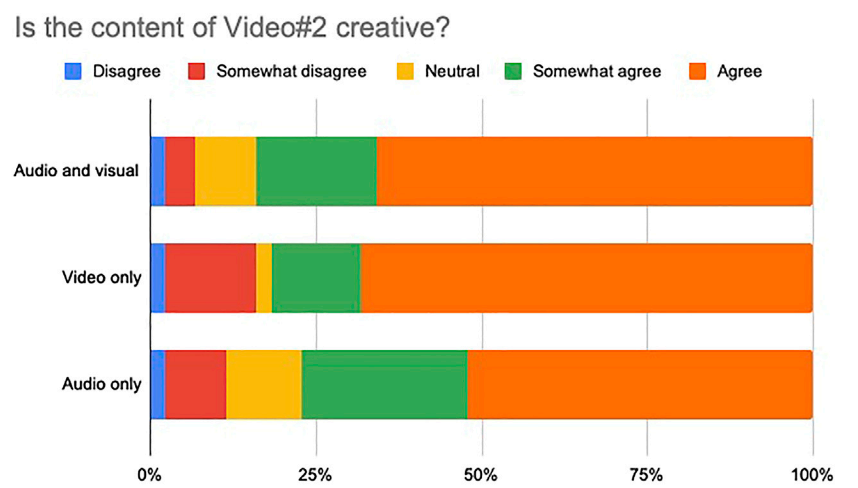


FIGURE 12 | Perceptual evaluation of Video #2. Forty-four participants responded the survey over the internet. All three type of contents: audiovisual, visual only and audio only, were judged to be creative by a great majority of participants.

6 DISCUSSION

We believe that both of the case studies that we have presented exhibits certain aspects of computational creativity. In particular, both networks can generate novel musical excerpts or chords different from the ones contained in their respective training sets, a clear signal of novelty, but that also make musical sense and can function very well in their musical context at the same time, a probable sign of value. In TimbreNet, the network is trained with only tertian triad chords, consisting of only three notes arranged by thirds, and exclusively major, minor, diminished or augmented chords. However, as **Figure 6** shows, the network can generate seventh chords, chords that are still tertian, but that contain four notes and that play a fundamental role in Western music, as their function within an harmonic context can be, for example, the dominant leading the way to the tonic, as in the case of the dominant seventh. If a chord-generating neural network trained with tertian triads exclusively can generate, after training, a dominant or minor-minor seventh chord, musical entities that the network had no clue they existed at all, does that make it a creative artificial intelligence? We believe that the answer must be yes, as the concept of a seventh chord is at the core of musical knowledge, and it is not trivial to derive from only regular tertian triads. In terms of value, it is interesting to notice that the network kept the configuration of new chords based on thirds, which makes musical sense. It could have simply generated lots of cluster chords, without any specific interval configuration, which would make them less coherent from a traditional Western harmonic point of view.

In general, a generative model is satisfactory if: 1) it can generate examples that appear to be drawn from the same distribution as the training dataset, a concept known as fidelity, and 2) the examples are suitably different from the examples shown during training, in other words, diversity (Naeem et al., 2020). In musical terms we can relate fidelity with adhesion to musical standards and diversity related to novelty and unexpectedness, all aspects of musical creativity (Daikoku et al., 2021). In the case of our experiments we found different degrees of achievement in fidelity and diversity depending on the number of dimensions of the latent space. For models with smaller latent space (3 or 4 dimensions) we found that the new chords were very similar to the chords in the dataset and no new different chords were generated, achieving fidelity but not diversity. For models with eight dimensions the chords were similar to the chords in the dataset but new chords with 4 or five notes were found. These new chords are suitably different from the training examples and they still have musical meaning and sense. We can say that this model achieved both fidelity and diversity. For models with bigger latent spaces (16 and 32 dimensions) new chords can be very different from those contained in the training set and they start loosing musical meaning and sense, achieving diversity, maximizing unexpectedness, but minimizing fidelity.

This network generates new chords when its latent space is sampled at coordinates that were not explicitly explored during training. It is indeed this sampling of uncharted territory that gives the possibility of something new and novel. This latent space is very similar to Boden's idea of a structured conceptual space, and this process of exploration is very congruent with the concept of exploratory creativity (Boden, 2004). This idea is also supported by Franceschelli and Musolesi (2021) and Basalla and Schneider (2020), who claim that VAEs are the best possible computational examples of exploratory creativity, as their main goal is to create a structured compressed space open to further exploration.

How is it possible that the network learned the concept of a seventh chord? We don't exactly know that at this point, but we propose that the fact that it learned that is a clear sign of creativity. One thing is being able to generate new audio based on chords, but a totally different thing is the ability to generate new chords, directly in audio, that fulfill a different tonal function with a different number of notes, but keeping its internal interval arrangement. In order to do that, TimbreNet must have learned the idea that a chord contains notes, that it can contain a variable number of them (even though it only saw tertian triads at training), and that these notes must be separated by major or minor thirds, in order to form a seventh chord. These, we insist, are not trivial concepts in music theory.

The fact that GANs possess a non-directly generated latent space, because the generator never sees real examples, implies that the sampling process in these kind of networks is from a conceptual space that could be indeed different from the original one, leading not only to exploratory creativity, but possibly also to transformational creativity (Franceschelli and Musolesi, 2021; Basalla and Schneider, 2020).

In effect, in StyleGAN Pianorolls, as it can be seen in **Figure 9** and heard in the audio examples, a network originally trained on images of human faces learned how to generate musical excerpts. Representations and features from the spatial domain of images were somehow transformed into musical ideas, a clear conceptual leap, and an example of transformational creativity (Boden, 2004). These musical ideas are also novel and posses musical value, a fact that strengthens the creative aspects of this network.

This network also exhibits self-evaluation, another critical aspect of creativity according to Moruzzi (2021). The discriminator does not allow "false" examples to survive, relying only on true examples to improve its performance. In a sense, these networks know "when to stop", without any need of external feedback.

The understanding of how the StyleGAN 2 model can learn to map different musical features, such as chords, scales and repeating motifs, to a new latent space to generate musical ideas that were not explicitly included in the training data, provides a new investigation opportunity to further explore how to use the disentangled space to get more control over the output of these models, so composers can use this as a tool for getting creative new ideas to overcome writers block for example.

In terms of the perceptual evaluation of the StyleGAN 2 model, we are aware that this is not a complete and rigorous perceptual evaluation of the creativity of our case studies. We also acknowledge that were are not comparing the results of these networks with those of a human counterpart, and that there is still some human intervention in the video production stage. Nevertheless, these results tend to confirm our hypothesis that these networks exhibits some traits of creativity, as their products were judged by a majority of our human subjects to be creative.

In summary, we have provided evidence that combined suggest that deep generative neural networks, can be, effectively, considered to be creative, or at least as creative as we consider humans are, based on our current understanding and knowledge on the topic. These networks generate valuable and novel outputs, and can conceptually leap, by using existing knowledge from a particular domain to generate knowledge in another domain. In the particular case of robot-generated music, these findings are particularly appealing and open a wide door for future creative possibilities.

7 CONCLUSION AND FUTURE WORK

The spectacular development of DL has not been alien to the world of the arts, as recent advances in generative models have made possible the creation of deep creative networks. As an example, we presented two case studies of our own: TimbreNet, a VAE network trained to generate audio-based musical chords, and StyleGAN Pianorolls, a GAN capable of creating short musical excerpts. We discussed and assessed these generative models in terms of their creativity and we show that they are capable of learning musical concepts that are not obvious based on the training data, they exhibit novelty, diversity, self-assessment, they can also produce conceptual leaps, and exploratory and transformational creativity. We have shown that these deep models, based on our current understanding of creativity in robots and machines, can be considered, in fact, creative.

In particular, we focused on the aspects of 1) novelty, in the sense that these models should produce something that is not expected, 2) value, by assessing whether novel outputs function well in a musical context, 3) exploratory creativity as they can represent complex ideas in a compact conceptual space, 4) self-assessent, in the sense that they do know when to stop, and 5) diversity transformational creativity and conceptual leaps, where one type of knowledge is used to produce a different kind.

For the purposes of this article, we used an evaluation strategy based both on the first and second strategies proposed by Jordanous (2019): first a creative-practitioner-type approach, i.e., perceptual evaluations by humans, and second, based on the assessment of the authors, as we were the creators of both models. In future work, we would like to incorporate more evaluation strategies in order to strengthen the argument that these networks can exhibit creative behavior, and a more complete subjective evaluation by humans. And also, we would like to dive in more depth into the exploration of the latent spaces of both modes, not only to show that these networks can be creative, but to understand why: what have they learned and how they acquired that knowledge and why it is that they can consider creative.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession numbers can be found below: https://zenodo.org/record/4740877 and https://zenodo.org/record/4747698.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This research was funded by the Dirección de Artes y Cultura, Vicerrectoría de Investigación from the Pontificia Universidad Católica de Chile. Also, this work is partially funded by Fondecyt grants #1161328 and #1191791 ANID, Government of Chile, as well as by the Millenium Institute Foundational Research on Data (IMFD).

REFERENCES

Andjelkovic, I., Parra, D., and O'Donovan, J. (2016). "Moodplay," in UMAP 2016 - Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization (Amsterdam, Netherlands: International Journal of Human-Computer Studies, Elsevier), 275–279. doi:10.1145/2930238.2930280

Andjelkovic, I., Parra, D., and O'Donovan, J. (2019). Moodplay: Interactive Music Recommendation Based on Artists' Mood Similarity. *Int. J. Human-Computer Stud.* 121, 142–159. doi:10.1016/j.ijhcs.2018.04.004

Basalla, M., and Schneider, J. (2020). Creativity of Deep Learning: Conceptualization and Assessment, 1–12.

Boden, M. A. (2009). Computer Models of Creativity. *AIMag* 30, 23–34. doi:10.1609/aimag.v30i3.2254

Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms*. 2nd editio edn. London and New York: Routledge.

Bretan, M., and Weinberg, G. (2016). A Survey of Robotic Musicianship. *Commun. ACM* 59, 100–109. doi:10.1145/2818994

Briot, J.-P., Hadjeres, G., and Pachet, F.-D. (2020). *Deep Learning Techniques for Music Generation*. Cham, Switzerland: Springer Nature. doi:10.1007/978-3-319-70163-9 Deep Learning Techniques for Music Generation.

Brown, R. T. (1989). "Creativity. What Are We to Measure?," in *Handbook of Creativity. Perspectives on Individual Differences*. Editors J. A. Glover, R. R. Ronning, and C. R. Reynolds (Boston, Mass: Springer). doi:10.1007/978-1-4757-5356-1–1

Cádiz, R. F. (2020). Creating Music with Fuzzy Logic. *Front. Artif. Intell.* 3, 1–20. doi:10.3389/frai.2020.00059

Carnovalini, F., and Rodà, A. (2020). Computational Creativity and Music Generation Systems: An Introduction to the State of the Art. *Front. Artif. Intell.* 3. doi:10.3389/frai.2020.00014

Carter, S., and Nielsen, M. (2017). Using Artificial Intelligence to Augment Human Intelligence. *Distill* 2, e9. doi:10.23915/distill.00009

Charniak, E. (2018). *Introduction to Deep Learning*. MIT Press.

Collins, D. (2007). Real-time Tracking of the Creative Music Composition Process. *Digital Creativity* 18 (4), 239–256. doi:10.1080/14626260701743234

Colton, S., Halskov, J., Ventura, D., Gouldstone, I., Cook, M., and Ferrer, B. P. (2015). *ICCC*, 189–196.The Painting Fool Sees! New Projects with the Automated Painter.

Cope, D. (1996). *Experiments in Musical Intelligence*. Madison, WI: A-R Editions.

[Dataset] Daikoku, T., Wiggins, G. A., and Nagai, Y. (2021). Statistical Properties of Musical Creativity: Roles of Hierarchy and Uncertainty in Statistical Learning. *Front. Neurosci.* 15. doi:10.3389/fnins.2021.640412

Deng, J., and Kwok, Y. K. (2016). "A Hybrid Gaussian-HMM-Deep-Learning Approach for Automatic Chord Estimation with Very Large Vocabulary, Editors Devaney, J., Mandel, M., Tzanetakis, G., Turnbull, D.," in Proceedings of the 17th International Society for Music Information Retrieval Conference, ISMIR 2016, August 7-11, 2016, New York City, USA, 812–818.

Dong, H.-W., Hsiao, W.-Y., Yang, L.-C., and Yang, Y.-H. (2018a). "Musegan: Multi-Track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment," in Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, New Orleans, LO.

Dong, H.-W., Hsiao, W.-Y., and Yang, Y.-H. (2018b). "Pypianoroll: Open Source Python Package for Handling Multitrack Pianorolls, Editors Flexer, A., Peeters, G., Urbano, J., Volk, A.," in 19th International Society for Music Information Retrieval Conference (ISMIR), September 23-27, 2018, Paris, France, 1–2.

Edmonds, E., Bilda, Z., and Muller, L. (2009). Artist, Evaluator and Curator: Three Viewpoints on Interactive Art, Evaluation and Audience Experience. *Digital Creativity* 20, 141–151. doi:10.1080/14626260903083579

Eigenfeldt, A., Burnett, A., and Pasquier, P. (2012). "Evaluating Musical Metacreation in a Live Performance Context, Editors Lou Maher, M., Hammond, K., Pease, A., Pérez y Pérez, R., Ventura, D., Wiggins, G.," in Proceedings of the 3rd International Conference on Computational Creativity, ICCC 2012, May 30 - June 1, 2012, Dublin, Ireland, 140–144.

Elgammal, A., Liu, B., Elhoseiny, M., and Mazzone, M. (2017). CAN: Creative Adversarial Networks, Generating "Art" by Learning about Styles and Deviating from Style Norms. ArXiv , 1–22.

Engel, J., Hantrakul, L., Gu, C., and Roberts, A. (2020). *ICLR*, 1–19.DDSP: Differentiable Digital Signal Processing.

Engel, J., Resnick, C., Roberts, A., Dieleman, S., Eck, D., Simonyan, K., et al. (2017). Neural Audio Synthesis of Musical Notes with Wavenet Autoencoders. arXiv preprint arXiv:1704.01279.

Franceschelli, G., and Musolesi, M. (2021). Creativity and Machine Learning: A Survey.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). *Advances in Neural Information Processing Systems*, 2672–2680. Generative Adversarial Nets.

Grace, K., and Maher, M. L. (2019). "Expectation-Based Models of Novelty for Evaluating Computational Creativity," in *Computational Creativity, Computational Synthesis and Creative Systems*. Editors T. Veale and F. A. Cardoso (Springer), 195–209. doi:10.1007/978-3-319-43610-4–9

Guzdial, M., and Riedl, M. (2019). "Combinets: Creativity via Recombination of Neural Networks, Editors Grace, K., Cook, M., Ventura, D., Lou, M.," in Proceedings of the 10th International Conference on Computational Creativity 2019, June 17-21, 2019, Charlotte, NC, 180–187.

Hadjeres, G., Pachet, F., and Nielsen, F. (2016). Deepbach: a Steerable Model for Bach Chorales Generation. arXiv preprint arXiv:1612.01010.

Hantrakul, L., Kondak, Z., and Weinberg, G. (2018). Practice Makes Perfect: Towards Learned Path Planning for Robotic Musicians Using Deep Reinforcement Learning. dl.acm.org doi:10.1145/3212721.3212839

Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C.-Z. A., Dieleman, S., et al. (2018). Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset, International Conference on Learning Representations, May 6-9, 2019. New Orleans, LO: ICLR.

Hennequin, R., Khlif, A., Voituret, F., and Moussallam, M. (2019). Spleeter: a Fast and State-Of-The Art Music Source Separation Tool with Pre-trained Models. ISMIR.

Hertzmann, A. (2018). Can Computers Create Art?. *Arts* 7, 18. doi:10.3390/arts7020018

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., et al. (2017). "B-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework," in 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings, April 24-26, 2017, Toulon, France, 1–22.

Humphrey, E. J., Cho, T., and Bello, J. P. (2012). "Learning a Robust Tonnetz-Space Transform for Automatic Chord Recognition," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, March 25-30, 2012, Kyoto, Japan, 453–456. doi:10.1109/ICASSP.2012.6287914

Jordanous, A. (2019). "Evaluating Evaluation: Assessing Progress and Practices in Computational Creativity Research," in Computational Creativity, *Computational* Synthesis And Creative Systems. Editors T. Veale and F. A. Cardoso (Springer), 211–236. doi:10.1007/978-3-319-43610-4–10

Kalin, J. (2018). *Generative Adversarial Networks Cookbook*. Birmingham, UK: Packt Publishing.

Karimi, P., Grace, K., Maher, M. L., and Davis, N. (2018). Evaluating Creativity in Computational Co-creative Systems. arXiv preprint arXiv:1807.09886

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., and Aila, T. (2020a). Training Generative Adversarial Networks with Limited Data

Karras, T., Laine, S., and Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networksin Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 14-19, 2020, 4401–4410. *Tech. Rep.*

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020b). "Analyzing and Improving the Image Quality of StyleGAN," in Proceedings of the CVPR.

Kim, H., Garrido, P., Tewari, A., Xu, W., Thies, J., Niessner, M., et al. (2018). Deep Video Portraits. *ACM Trans. Graphics*, 37 (4), 1–14. doi:10.1145/3197517.3201283

Kingma, D. P., and Welling, M. (2014). "Auto-encoding Variational Bayes," in 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings, April 14-16 2014, Banff, Canada, 1–14.

Korzeniowski, F., and Widmer, G. (2016). "Feature Learning for Chord Recognition: The Deep Chroma Extractor, Editors Devaney, J., Mandel, M. I., Turnbull, D., Tzanetakis, G. ," in Proceedings of the 17th International Society for Music Information Retrieval Conference, ISMIR 2016, August 7-11, 2016, New York City, USA, 37–43.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* doi:10.1061/(ASCE)GT.1943-5606.0001284

Miranda, E. R., and Witkowski, M. (2005). "Musical Composition by Autonomous Robots: A Case Study with AIBO, Editors Nehmzow, U., Melhuish, C., Tikhanoff, V.," in Proceedings of Toward Autonomous Robotic Systems (TAROS), September 12-14 2005, London, UK.

Moruzzi, C. (2018). "Creative AI: Music Composition Programs as an Extension of the Composer's Mind,". *Studies in Applied Philosophy, Epistemology and Rational Ethics*. Editor V. C. Muller (Springer), 44, 69–72. doi:10.1007/978-3-319-96448-5–8

Moruzzi, C. (2021). Measuring Creativity: an Account of Natural and Artificial Creativity. *Eur. J. Philos. Sci.* 11, 1–20. doi:10.1007/s13194-020-00313-w

Mumford, M., and Ventura, D. (2015). "The Man behind the Curtain: Overcoming Skepticism about Creative Computing, Editors Toivonen, H., Colton, S., Cook, M., Ventura, D.," in Proceedings of the Sixth International Conference on Computational Creativity, June 29–July 2, 2015, Park City, Utah.

Naeem, M. F., Oh, S. J., Uh, Y., Choi, Y., and Yoo, J. (2020). Reliable Fidelity and Diversity Metrics for Generative Models. *International Conference on Machine Learning*. Vienna, Austria.

Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., et al. (2016). Wavenet: A Generative Model for Raw Audio. arXiv preprint arXiv:1609.03499.

Park, Y. (2019). Can Artworks by Artificial Intelligence Be Artworks?. *AM. J. Art Media Stud.* 113. doi:10.25038/am.v0i20.332

Raffel, C. (2016). Learning-Based Methods for Comparing Sequences, with Applications to Audio-To-MIDI Alignment and Matching. Ph.D. thesis. Columbia University.

Ranganath, R., Gerrish, S., and Blei, D. M. (2014). Black Box Variational Inference. *J. Machine Learn. Res.* 33, 814–822.

Ritchie, G. (2019). "The Evaluation of Creative Systems," in *Computational Creativity, Computational Synthesis and Creative Systems,*. Editors T. Veale and F. A. Cardoso (Springer), 159–194. doi:10.1007/978-3-319-43610-4–8

Roberts, A., Engel, J., and Eck, D. (2017). Hierarchical Variational Autoencoders for Music. *NIPS Workshop on Machine Learning for Creativity and Design*, 3.

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., and Eck, D. (2018). A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music. arXiv preprint arXiv:1803.05428.

Rowe, R. (2004). *Machine Musicianship*. Cambridge, MA: The MIT Press. doi:10.7551/mitpress/4361.003.0002

Sawyer, R. K. (2006). *Explaining Creativity. The Science of Human Innovation*. Oxford University Press.

Sternberg, R. J., and Sternberg, K. (2012). *Cognitive Psychology*. sixth edit edn. Belmont, CA: Wadsworth Cengage Learning.

Sturm, B. L., Ben-Tal, O., Monaghan, U., Collins, N., Herremans, D., Chew, E., et al. (2018). Machine Learning Research that Matters for Music Creation: A Case Study. *J. New Music Res.* 48, 36–55. doi:10.1080/09298215.2018.1515233

Sturm, B. L., Santos, J. F., Ben-Tal, O., and Korshunova, I. (2016). Music Transcription Modelling and Composition Using Deep Learning. *arXiv preprint arXiv:1604*.08723.

Weber, A., Alegre, L. N., Tørresen, J., and da Silva, B. C. (2019). Parameterized Melody Generation with Autoencoders and Temporally-Consistent Noise, Proceedings of the New Interfaces for Musical Expression Conference, June 3-6, 2019. Porto Alegre, Brazil: NIME, 174–179.

Wyse, L. (2019). "Mechanisms of Artistic Creativity in Deep Learning Neural Networks, Editors Grace, K., Cook, M., Ventura, D., Lou, M.," in Proceedings of the 10th International Conference on Computational Creativity 2019, June 17-21, 2019, Charlotte, NC.

Yamshchikov, I. P., and Tikhonov, A. (2020). Music Generation with Variational Recurrent Autoencoder Supported by History. *SN Appl. Sci.* 2, 1–7. doi:10.1007/s42452-020-03715-w

Yang, L.-C., Chou, S.-Y., and Yang, Y.-H. (2017). MidiNet: A Convolutional Generative Adversarial Network for Symbolic-Domain Music Generation. arXiv preprint arXiv:1703.10847.

Yang, L. C., and Lerch, A. (2020). On the Evaluation of Generative Models in Music. *Neural Comput. Appl.* 32, 4773–4784. doi:10.1007/s00521-018-3849-7

Zhou, X., and Lerch, A. (2015). "Chord Detection Using Deep Learning, Editors Müller, M., Wiering, F.," in Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015, October 26-30, 2015, Málaga, Spain, 52–58.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer IA declared a past co-authorship with one of the authors (DP) to the handling Editor.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 13 September 2021
doi: 10.3389/frobt.2021.673730

Social Robots as Creativity Eliciting Agents

Safinah Ali*, Nisha Devasia, Hae Won Park and Cynthia Breazeal

MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, United States

Can robots help children be more creative? In this work, we posit social robots as *creativity support tools* for children in collaborative interactions. Children learn creative expressions and behaviors through social interactions with others during playful and collaborative tasks, and socially emulate their peers' and teachers' creativity. Social robots have a unique ability to engage in social and emotional interactions with children that can be leveraged to foster creative expression. We focus on two types of social interactions: *creativity demonstration*, where the robot exhibits creative behaviors, and *creativity scaffolding*, where the robot poses challenges, suggests ideas, provides positive reinforcement, and asks questions to scaffold children's creativity. We situate our research in three playful and collaborative tasks - the Droodle Creativity game (that affords verbal creativity), the MagicDraw game (that affords figural creativity), and the WeDo construction task (that affords constructional creativity), that children play with Jibo, a social robot. To evaluate the efficacy of the robot's social behaviors in enhancing creative behavior and expression in children, we ran three randomized controlled trials with 169 children in the 5–10 yr old age group. In the first two tasks, the robot exhibited creativity demonstration behaviors. We found that children who interacted with the robot exhibiting high verbal creativity in the Droodle game and high figural creativity in the MagicDraw game also exhibited significantly higher creativity than a control group of participants who interacted with a robot that did not express creativity ($p < 0.05^*$). In the WeDo construction task, children who interacted with the robot that expressed creative scaffolding behaviors (asking reflective questions, generating ideas and challenges, and providing positive reinforcement) demonstrated higher creativity than participants in the control group by expressing a greater number of ideas, more original ideas, and more varied use of available materials ($p < 0.05^*$). We found that both creativity demonstration and creativity scaffolding can be leveraged as social mechanisms for eliciting creativity in children using a social robot. From our findings, we suggest design guidelines for pedagogical tools and social agent interactions to better support children's creativity.

Keywords: creativity, social robots, emulation, scaffolding, creativity support tools, child-robot interaction, collaboration

OPEN ACCESS

Edited by:
Patricia Alves-Oliveira, University of Washington, United States

Reviewed by:
Kerstin Fischer, University of Southern Denmark, Denmark
Sangseok You, HEC Paris, France

Correspondence:
Safinah Ali
safinah@media.mit.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *28 February 2021*
Accepted: *11 August 2021*
Published: *13 September 2021*

Citation:
Ali S, Devasia N, Park HW and Breazeal C (2021) Social Robots as Creativity Eliciting Agents. Front. Robot. AI 8:673730. doi: 10.3389/frobt.2021.673730

INTRODUCTION

Children's creativity–their ability to generate novel, surprising, and valuable ideas–is known to contribute to their learning outcomes, personal growth, and well-being. Creativity facilitates children's problem solving, adaptability, self-expression and health (Carterette et al., 1994). Even though the benefits of creativity are widely recognized, classrooms are not able to sufficiently support children's creative development. Gardner and Art, 1982 posited that children start to show creative abilities as early as preschool, and Smith and Carlsson (1983) found that the level of developmental maturity necessary for creative expression occurs around 5–6 yr of age. However, as they enter elementary school, children's creativity slumps, especially around the fourth grade (Torrance, 1966, 1968; Claxton, 2005). As school curricula become more structured, children lose the aspect of creative play that is a significant part of kindergarten. To be successful in our AI-powered world, where mechanical and repetitive jobs are becoming automated, we must empower children to generate new artifacts and solve complex problems, which will require imaginative and novel thought.

In classrooms, social interaction plays a key role in children's creative growth. Children learn creativity from teachers and peers who act as models for creative expression. They can scaffold children's creativity through social interactions such as collaborating, posing challenges, asking questions, providing positive reinforcement, and generating ideas. Participating in collaborative tasks is one of the most effective external influencers of creativity (Kafai et al., 1995). In addition, classrooms today see increasing numbers of digital pedagogical tools and learning aids that have proven to be beneficial for cognitive learning due to their ability to personalize instruction for every student. Games and play-based learning approaches have been successful in fostering creative expression in children (Henriksen, 2006; Bowman et al., 2015). However, most educational technologies do not have the ability to foster social interaction with students. Exceptions include socially interactive AI agents such as conversational agents and social robots, which are highly effective in promoting learning and engagement (Chen et al., 2020). Previous work has demonstrated how social robots can influence children's learning behaviors such as curiosity, growth mindset, and empathy through social emulation (Gordon et al., 2015; Park et al., 2017). Notably, the presence of a social robot can also affect adults' creativity (Kahn et al., 2016; Alves-Oliveira et al., 2019). In this work, we explore how social robots can foster creativity in young children through collaborative, playful interactions.

Social robots are increasingly being used as learning peers and tutors (Adamson et al., 2021), and their unique ability to socially interact with children while being co-located has situated them well as creativity support tools (CST). Previous work has shown that robots are more effective than other mediums used for CSTs, such as screen-based interfaces (Kahn et al., 2016; Ali, 2019). We do not seek to compare robots to other mediums of CSTs in this work; rather, we intend to demonstrate the efficacy of creativity stimulating interactions designed for social robots. Furthermore, while social interaction is not a prerequisite for creativity, creativity literature informs us that social interactions with peers and tutors can foster creativity in children. Previous work has shown how situating the robot as a collaborative peer that offers ideas or helps with the creative process have benefitted creative expression (Louis and Peter, 2015; Rond et al., 2019). In this work, we explore whether a social robot's capability for social interaction patterns can stimulate children's creativity, while acknowledging that there are other stimulants of creativity. Learning from the effect of sociality on creativity in classrooms, we explore whether the effect can be replicated in pedagogical tools, specifically social robots. We suggest two interaction patterns in which intelligent embodied agents can help children think more creatively: *1*) creativity modeling, where the social robot models or demonstrates desired creative behaviors, and *2*) creativity scaffolding, where the robot offers scaffolding to the child in the form of asking reflective questions, validating novel ideas, and engaging in creative conflict. The robot used in this work is Jibo -- a child-friendly, tabletop, socially expressive robot.

We position our research in three playful and collaborative tasks, where children and the robot collaborate to create artifacts. These one-on-one interactions afford different forms of creative expression.

In the first two tasks, outlined in our previous research (Ali et al., 2019; Ali et al., 2020), we designed the behavior of the robot to artificially emulate human creativity:

1. The Droodle Creativity game, where children and the robot generate humorous titles for abstract images to express verbal creativity.
2. The Magic Draw game, where children and the robot co-create drawings on a tablet screen to express figural creativity.

In this work, we introduce the third task, where the robot scaffolds children's creative thinking by asking questions, validating novel ideas, and engaging in creative conflict:

3. The WeDo Construction activity, where children and the robot co-create WeDo LEGO models to encourage constructional creativity.

We outline previous research, where we demonstrated that children can emulate a social robotic peer's creative expression during collaborative gameplay. We ran two randomized controlled trials with 126 children: verbal creativity (n = 48) and figural creativity (n = 78) in the 5–10 yr old age group. Participants in the intervention group interacted with the robot exhibiting creative behaviors and participants in the control group interacted with the robot that did not exhibit these behaviors. We observed that children who interacted with the robot exhibiting high verbal creativity (in the Droodle Creativity game) and high figural creativity (in the MagicDraw game) exhibited higher verbal and figural creativity themselves. We then introduce our third study, where the robot offers creativity scaffolding behaviors in the WeDo Construction task. We ran a randomized controlled trial with 42 students in the 6–10 yr old age group. We observed that when the robot offered creativity scaffolding in the construction task, children expressed higher creativity.

In sum, we provide consistent evidence that the performance of creativity inducing behaviors by social robots can foster creativity in young children. Further, in all three studies, children in the high creative robot conditions perceived the robot as more creative and fun as compared to the low creative conditions. By showing that a social robot can successfully foster different kinds of creative expression, we are able to articulate more generalized social mechanisms that can be leveraged to support creativity in children. We contribute novel design guidelines and new methods for designing interactions with social agents that aim to promote creative thinking. With new developments in generative modeling techniques, robots can participate in several co-creative tasks with children, and can be leveraged as creativity support tools in a wide range of creative activities. We discuss implications for the field of HRI, digital creativity support tools, co-creative agents, and transformative games.

BACKGROUND

Creativity

Creativity is often referred to as the ability to generate artifacts or ideas that are both novel and appropriate to the problem at hand (Carterette et al., 1994). Novelty is the ability to generate ideas that are different from one's own ideas, and different from the group's ideas. The appropriateness of a solution refers to solving problems using the least amount of time and resources. The definition of creativity has evolved from a function of the individual to an interaction between aptitude, environment, and process by which an individual produces a tangible product (Plucker et al., 2004). Depending on the nature of the task and the medium of creative expression, creativity presents itself in different forms; for example, figural creativity (drawing, painting, sketching) and verbal creativity (writing, storytelling, composition, discourse) (Guilford, 1957). In addition, construction (building, tinkering) is described as a form of creativity in which students can draw their own conclusions through creative experimentation and the creation of artifacts (Harel and Papert, 1991). A common means for fostering creative learning in classrooms is through construction and maker based activities, which we refer to as constructional creativity in this work.

Early creativity researchers defined creativity as the embodiment of thought in the form of external behavior, consisting of three characteristics: fluency, flexibility, and originality (Guilford, 1950). Fluency refers to the ability to generate several ideas, flexibility refers to the variation in themes between several generated ideas, and originality refers to the novelty of the ideas generated in comparison to those of the group's. For the purpose of this work, we define the ability to generate ideas with greater fluency, flexibility and originality as creative thinking. Metrics of fluency, flexibility and originality are dependent on the choice of tasks made for each type of creativity. We also take divergent thinking into account as a component of creativity, and categorize activities that involve the creation of artifacts as activities that afford creativity.

Extrinsic Factors Influencing Creativity

Researchers have identified several factors that may serve as "situational influences" of creativity: freedom, autonomy, good role models and resources, encouragement for originality, little criticism, and "norms in which innovation is prized and failure not fatal" (Amabile and Gryskiewicz, 1989; Witt and Beorkem, 1989). In our work, we utilize the following factors to design effective creative interactions.

Emulation

Emulation is described as "[when] children achieve common goals to those modelled, but do so by using idiosyncratic means that were never observed" (Bornstein and Bruner, 2014). Indeed, children are predisposed to social emulation (Yando et al., 1978), and learn from other creators in their environments, such as teachers and classmates, through mechanisms of social emulation (Whiten et al., 2009). Within classroom settings, researchers have suggested that the traditional educational model, which emphasizes rote problem solving, can be overcome by providing students with more diverse models of creativity to emulate (Root-Bernstein and Root-Bernstein, 2017). Social emulation may even be at the heart of innovation itself; one study showed that in a tower building task, children performed poorly at the task independently, but after observing one or two models building a tower, they were able to emulate the demonstrated elements and spontaneously recombine them, producing a novel tower of an optimal height (Subiaul and Stanton, 2020).

Social Interactions in the Classroom

The importance of the social environment to creativity is well researched (Kaufman and Sternberg, 2010). For children, the primary social environment is the classroom. Several factors that influence creativity, such as emulation, play, and collaboration, are heavily integrated into early education classroom curricula (Halverson and Sheridan, 2014; Kafai et al., 1995). Question-asking during creative activities stimulates their creativity (Zheng et al., 2013), and creative learning research outlines how game-based learning environments must facilitate reflective thinking (generating ideas and evaluating them) in order to foster creativity (Henriksen, 2006; Bowman, et al., 2015).

Collaboration

Creativity has typically been understood as an individualistic pursuit. However, it is now widely accepted that creativity stems from the confluence of diverse perspectives and ideas, and that the nature of collaboration stimulates creative problem solving (Kafai et al., 1995). For children in particular, several studies emphasize the importance of friendship and peership in fostering effective creative collaboration (MacDonald et al., 2000; Miell and MacDonald, 2000; Bass et al., 2008). The studies described in this work utilize a social robot as a peer in order to stimulate creative collaboration with children.

Play

Play-based learning tools and game-like activities have been repeatedly shown to promote creativity (Henriksen, 2006; Garaigordobil, 2006; Baggerly, 1999; Dansky, 1980; Howard-Jones et al., 2002; Mellou, 1995; Russ, 2003; Berretta and Privette, 1990). They are effective for teaching concepts to children since their entertainment value ensures higher engagement levels. Furthermore, several behaviors that constitute creativity can be promoted *via* gameplay behaviors, such as developing multiple solutions to a problem, generating novel and appropriate solutions, metacognition, question-asking, and cross-contextual thinking (Henriksen, 2006). Games designed to specifically alter players' behaviors, attitudes, or knowledge during and after play are known as *transformational games*, which can be used as tools to support meaningful learning. Digital games in particular provide players with the opportunity to find many creative solutions within a singular play space (Bowman et al., 2015), especially in the case of well-known sandbox games such as Minecraft (Duncan, 2011). In our work, we utilize transformational digital games with game mechanics that allow for creative expression and creative problem-solving.

Creativity Support Tools

Given the many extrinsic influencers of creativity, it is no surprise that HCI researchers have attempted to engineer creativity support tools (CSTs) "that empower users to be not only more productive but also more innovative" (Shneiderman et al., 2006). Since the framework's proposal (Shneiderman, 2002), researchers have developed a wide range of CSTs. The vast majority of CSTs were built for digital devices, with the most common being a laptop or a personal computer (Frich et al., 2019). Previous work has demonstrated how creativity is also facilitate through analog toolkits such as Scratch Coding Cards (Scratch Team, 2017) and robotic construction kits such as Lego Mindstorms, Popbots, and Cozmo (Anki, 1999; Williams et al., 2019).

Social Robots as CSTs

Despite creativity's social nature, little work has been conducted on the benefit of utilizing social agents as CSTs. Previous work has demonstrated how verbal and non-verbal social robot behavior can serve to engage adults in a creative activity for longer and aid their own creative ideas (Kahn et al., 2016; Alves-Oliveira et al., 2019), and children will emulate a social robot's expressed verbal and figural creativity, resulting in a higher level of creative expression (Ali et al., 2019; Ali et al., 2020). Alves-Oliveira et al. (2020) demonstrated how interacting with the robotic system YOLO when it displayed social and creative behaviors simulated children's creative abilities. Robots with light patterns have also benefited children's storytelling experiences (Ligthart et al., 2020). Another study showed how people spent more time creating music with drums while collaborating with the Mortimer robot (Louis and Peter, 2015). Rond et al. (2019) found that adult improve performers viewed a simple robot as a supportive teammate who positively inspired the scene's direction. A majority of previous work utilized social robots as peers or partners. However, all mentioned works are specific to one creative task.

Social robotic agents are proven to be effective learning companions (Belpaeme et al., 2018), and children form relationships with them through social interactions (Westlund et al., 2018). There lies a unique opportunity in being able leverage these social and interactive agents as creativity fostering mechanisms for children. In this work, we demonstrate the efficacy of social robots as CSTs through three creative tasks that focus on three different kinds of creativity: verbal, figural and constructional. Further, we utilize game-based interactions since play is known to benefit creativity and it helps situate the robot as a collaborative playful peer. Like previous work, we situated the robot as a collaborative peer (Rond et al., 2019). Similar to Ali et al. (2019) and (Ali et al., 2021), we utilize the robot's creativity demonstration as a creativity eliciting mechanism. Similar to Alves-Oliveira et al. (2019) and Kahn et al. (2016), we made use of the robot's social verbal and non-verbal interactions during child-robot interaction. Through the three game interaction, and the robot assuming different social behaviors, we studied how creativity demonstration and creativity scaffolding through social interactions benefited children's creative expression. We suggest interaction patterns of social robots specific to a computational learning setting that aim to foster creativity (described as Creativity Scaffolding interactions) that are generalizable to other creative tasks for children. Through this work, we aim to contribute to the literature of using social interactive agents as creativity support tools, through both their social interactions scaffolding children's creativity, and their creativity demonstrations acting as a model for children to emulate.

ROBOT PLATFORMS

For the three creativity activities, we used Jibo, a socially embodied robot, as our robotic platform (Jibo, 2015). Jibo is an expressive tabletop social robot (**Figure 1**) that can speak, respond to children's speech, track faces, gaze toward the person, attend to sound and movement in its environment, and physically express emotion through its display and three degree-of-freedom body. Jibo can communicate with an Android tablet that serves as a shared drawing surface or to display information for the child. Prior to each study, the experimenters discussed with children that Jibo uses WiFi to see, talk, draw, and interact with objects displayed on the tablet, and that it doesn't need physical hands to do so. Such discussion is important to building a believable experience for children that the robot knows what is displayed on the tablet and can draw on the tablet, too.

During the interaction, Jibo provides explanations and encouragement to the child, and expresses joy, curiosity, and pride. Between the high creative (C+) and low creative (C−) robot conditions, we carefully controlled the amount of verbal and non-verbal robot behaviors; however, the robot's speech differed based on the study condition. In the C+ condition, the robot exhibited greater creativity; in the C− condition, it demonstrated substantially lower creativity.

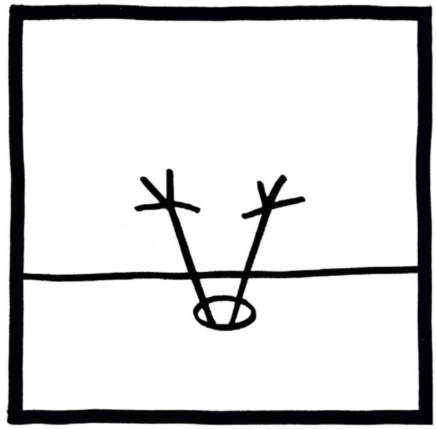


FIGURE 1 | Interaction Scene. **(A)** A child is playing the Droodle Creativity Game on an Android Tablet with social robot Jibo. **(B)** Example of a Droodle Image. 10 Droodles were used in the Droodle Creativity Game (five per player).

We developed three game-based and interaction child-robot tasks which we introduce in the following sections.

EXPERIMENT 1: DROODLE CREATIVITY GAME

Previous work in HRI demonstrated how children emulate robots' learning behaviors such as curiosity and growth mindset (Gordon et al., 2015; Park et al., 2017). Motivated from this literature, we explored whether this social emulation phenomenon extends to verbal creativity. Verbal creativity, defined as the ability to create verbal artifacts such as stories, prose and poetry has three indicators: fluency, or the ability to produce a large number of ideas; flexibility, or the aptitude for changing from one approach to another or from one line of thinking to another; and originality, or the capacity for bringing new ideas or solutions that are far from obvious, common, or established. Development of verbal creativity in children is pivotal to their learning, writing and thinking skills and helps them reflect their feelings, emotions, opinions, reactions, and notions to others (Shorofat, 2007; Rababah et al., 2017). In this work we explore whether a social robot's creative verbal expression is emulated by children, for which we created the Droodle Creativity game (Ali et al., 2019), inspired by the Droodle Creativity Task, a verbal creativity task that draws upon people's ability to creatively use language to describe an abstract image or figure known as a *droodle* (Kahn et al., 2005). The Droodle Creativity Task has been previously validated as a means to measure people's verbal creativity, and is based on the cartoon book Droodles by Roger Price (1982), thus making it appropriate for a children's game. The Droodle game encourages children to think creatively, express their thoughts and encourage humor.

Game Design

In the Droodle Creativity Game, two players take turns to generate Droodle titles (**Figure 1**). The active player is presented with droodles on a tablet screen and they come up with droodle title(s) in 30 s. Then the turn shifts to the other player until each player has played five turns each. The Droodle Task coding system, developed by Kahn et al. (2005), provides a metric for ranking the titles as "non-droodle," "low-," "medium-," or "high-droodle" based on the participant's initial reaction, pattern matching to the image in question, and reasoning for providing such an answer. Droodles used in our study were taken from *Droodles: The Classic Collection* (Price and Lovka, 2000) which also includes a library of droodle titles.

Interaction Scenario

When the interaction starts, Jibo explains to the child how the Droodle task works and engages the child in a practice round. When the child generates a creative droodle title, the robot praises the child by using phrases such as, "Great job," "I would not have thought of that," or "You are doing great." When the robot is "thinking" of a Droodle idea, it expresses curiosity through questioning sounds, swaying movements, and looking upwards.

Experiment Design

Participants

We recruited 51 subjects in the 5–10 yr age range as a part of the Somerville after-school activities program at the public schools in Somerville, MA. All students had basic knowledge of robotics and artificial intelligence taught to them as a part of another module of the after-school program. Three students were excluded due to technological malfunctions or a rudimentary understanding of English.

All participants and their guardians signed an informed assent and consent form to participate in the study and permit us to collect demographic, assessment, audio and video data. The recruitment materials, study protocol, and data collection protocol were reviewed and approved by the Institute Research Board at Massachusetts Institute of Technology.

Pre-test

All students completed the Torrance Test of Creative Thinking (TTCT) assessment as a part of the pre-test activity (Torrance,

TABLE 1 | A gameplay comparison of the high creative and low creative study conditions.

	Fluency	Novelty	Value
High creative robot (C+)	Robot generated four to five ideas per Droodle	Robot explored three or more different themes	Robot picks Droodle titles that are tagged medium/high in creativity
Low creative robot (C–)	Robot generated one to two ideas per Droodle	Robot explored one to three different themes	Robot picks Droodle titles that are tagged low/medium in creativity (e.g. the literal description of an image)

1968). The TTCT is a paper-based evaluation that consists of two sets of assessment activities: a verbal creativity test and a figural creativity test. The purpose of conducting the TTCT before the study was to drive a quasi-random assignment into groups such that their creativity scores are counterbalanced across the two conditions, described in the following section.

Study Conditions

Forty-eight participants were divided into two study condition groups, one that interacted with the high creative robot (C+) and one that interacted with the low creative robot (C–). The groups were divided such that the participants in the two groups were balanced in terms of their mean and standard deviations of TTCT scores (C+: 42.16 ± 7.17; C–: 40.66 ± 6.01), age (C+: 7.78 ± 1.92; C–: 8.38 ± 1.85), and gender (C+: F = 9, M = 15; C–: F = 13, M = 11).

The robot exhibits high or low creativity during gameplay depending on the study condition (**Table 1**). We use Boden's framework of creativity to design creative behaviors in gameplay (Boden, 2004), and Kahn et al.'s Droodle Task Coding system (Kahn et al., 2005) to determine the creativity of Droodle titles.

Hypotheses

H1: Participants interacting with the high creative robot (C+) generate a larger *number* of ideas than participants interacting with the low creative robot (C–).

H2: Participants interacting with the high creative robot (C+) explore more *themes* of ideas than participants interacting with the low creative robot (C–).

H3: Participants interacting with the high creative robot (C+) generate more *creative* ideas than participants interacting with the low creative condition (C–).

Data Collection and Measures

Children's speech and video data was recorded. We used Google Cloud's Speech API (Google, 2020), as well as manual transcribing by three researchers blind to the study to transcribe children's phrases. We used the TTCT to assess children's verbal and figural creativity prior to all study interactions, and to divide them into balanced study groups.

We measured participants' creativity in three parts:

- *Fluency.* The number of ideas that the participants generated.
- *Novelty.* The number of unique themes explored through the ideas. Each idea is associated with theme tags, which include all concepts and keywords included in the idea.
- *Value.* The droodle creativity scores of the ideas generated. Droodles are graded on a scale from 0 to 3, mapping to non-droodle, low-droodle, medium-droodle, and high-droodle respectively (Kahn et al., 2005).

For instance, one participant came up with the following ideas for the droodle image in round 1 (**Figure 1**): "*It's peppa pig*"; "*It's peppa pig's hands*"; and "*It's frog hands*." This would be analyzed as: Number of ideas (fluency) = 3; Unique themes (novelty) = "*peppa pig*," "*hands*," "*frogs*"; Droodle scores (value): 2, 3, 2.

Results

We calculated numerical values for each of the three creativity measures, then further determined the mean and standard deviation of the *Novelty* and *Value* scores for every Droodle image for each participant. For instance, if a participant generated three ideas for Droodle #1, the *Novelty* and *Value* would be the mean score of the three individual *Novelty* and *Value* scores. We then conducted unpaired T-tests between the high creative and low creative study participants to determine any between group differences for each of the three metrics.

H1: Participants interacting with the high creative robot (C+) generate a larger *number* of ideas than participants interacting with the low creative robot (C–).

To test our first hypothesis, we analyzed the number of ideas generated by the participants in the two study conditions. We observed that participants who interacted with the robot expressing high levels of creativity (C+) generated significantly more ideas ($t(29) = 1.699$, $p < 0.01$**) compared to the participants who interacted with the robot expressing low levels of creativity (C–) (**Table 2**).

H2: Participants interacting with the high creative robot (C+) explore more *themes* of ideas than participants interacting with the low creative robot (C–).

To understand the novelty of the themes that participants generated, we used the Rapid Automatic Keyword Extraction algorithm (Rake NLTK), a natural language processing library, to analyze the themes explored in each title (Mumford, 2001). We observed that participants in the C+ condition explored significantly more overall unique themes ($t(29) = 1.699$, $p < 0.01$**) as compared to the participants who interacted with the robot expressing low levels of creativity (**Table 2**).

H3: Participants interacting with the high creative robot (C+) generate more *creative* ideas than participants interacting with the low creative condition (C–).

TABLE 2 | Droodle Creativity game *t*-test results per condition for each study hypothesis.

SG	Ideas generated (H1)	Themes explored (H2)	Creativity scores per Droodle (H3)
C+ (*n* = 24)	3.325 ± 1.16	4.983 ± 1.25	1.73 ± 0.21
C− (*n* = 24)	2.417 ± 0.96	3.842 ± 1.66	1.532 ± 0.25
p	$t(29) = 1.699$, $p = 0.006$	$t(29) = 1.699$, $p = 0.010$	$t(29) = 1.699$, $p = <0.015$

Participants in the C+ condition generated more ideas, explored more themes, and overall received higher creativity scores than participants in the C− condition.

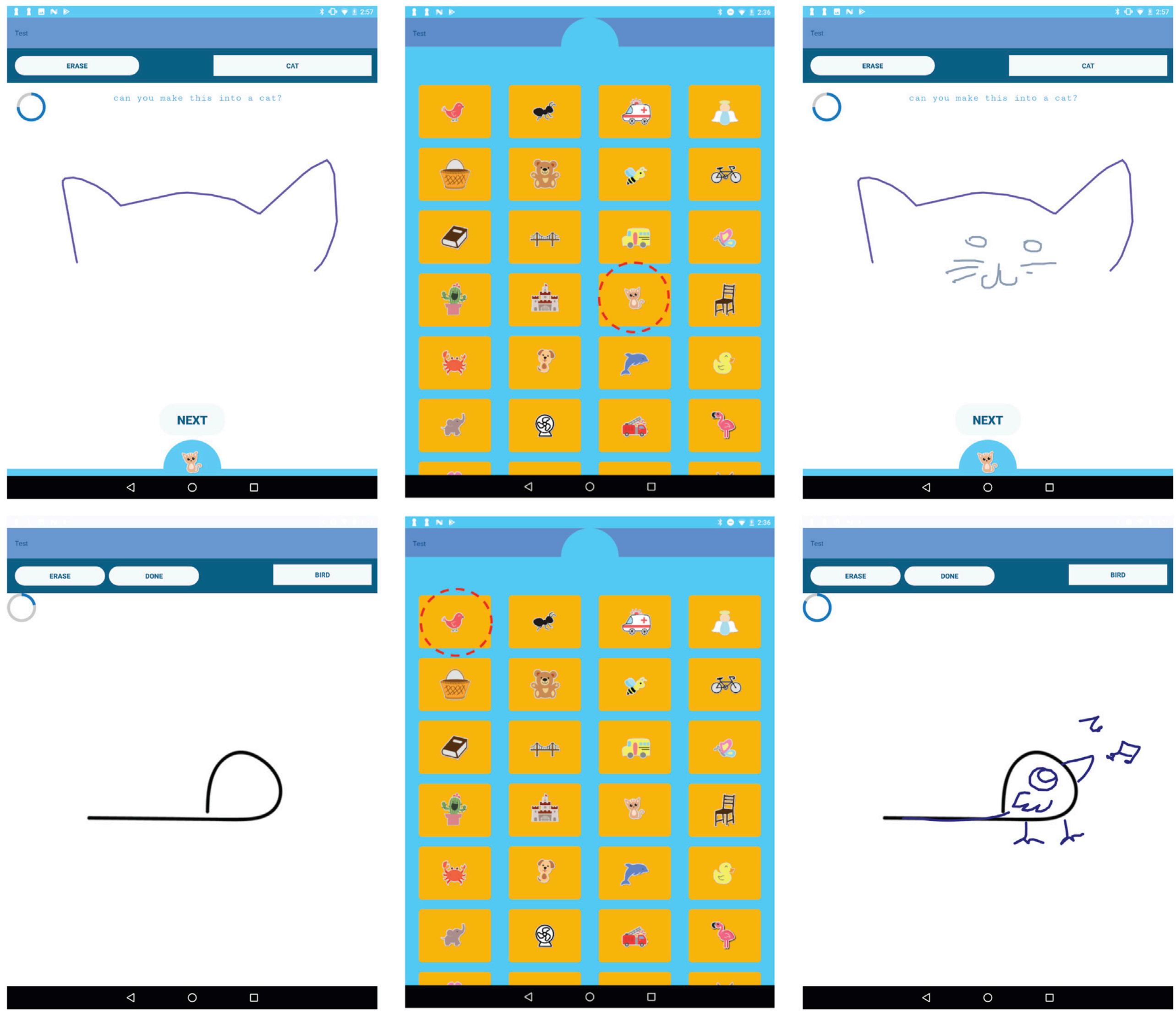


FIGURE 2 | Interface screenshots of the Magic Draw game explaining the child-robot gameplay. **(A)** The child draws a starting prompt (cat ears). **(B)** The child selects a target category (cat). **(C)** The robot tries to convert the starting prompt (circle) into the target category (cat). **(D)** Players switch turns and the robot gives the child a starting prompt. **(E)** The robot selects the category (bird). **(F)** The child converts the starting prompt to a bird. (order: left to right starting top-left).

Three coders blind to the study conditions were trained using the Droodle creativity coding scheme. They then coded all Droodle titles generated by the participants as "non-," "low-," "medium-" and "high-droodle." To determine inter-rater reliability between researchers, Cohen's kappa (Hallgren, 2012) was calculated using 67% of the coded transcripts coded independently by a team member after an initial coding by other two coders. Cohen's kappa was 0.82, which is within the

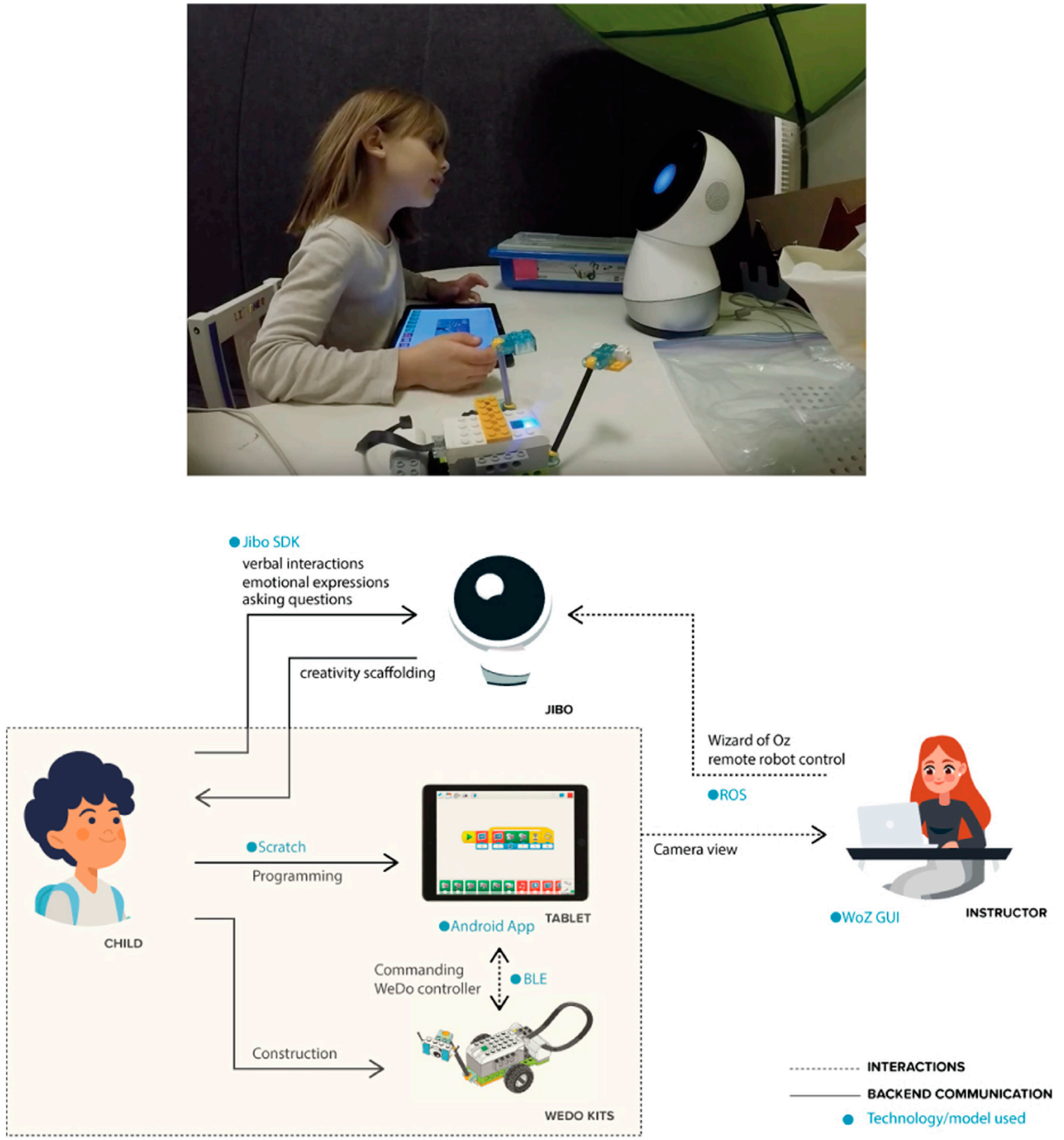


FIGURE 3 | (top) **(A)** Child constructing models with Jibo. (bottom) **(B)** Model of interaction for the WeDo construction game where the robot provides creativity scaffolding.

range for substantial agreement considered acceptable for inter-rater reliability.

An overall analysis of creativity scores for every idea revealed that participants in the creative condition scored significantly higher in creativity score per title than participants in the low creative condition ($t(29) = 1.699$, $p < 0.01$**) (**Figure 2**).

EXPERIMENT 2: MAGICDRAW

Similar to verbal creativity, we aimed to explore whether children also emulate a robots' expressed figural creativity. Generative models such as GANs make it possible for AI models to generate creative drawings (Ha and Eck, 2017; Ge et al., 2020).

Collaboration with both humans and digital interfaces has also been known to benefit figural creativity (Kim et al., 2016). Collaborative digital CSTs have been constantly evolving to support the creativity needs of people, and have the ability to contribute to the drawing itself. In human-human creative collaborations, creators can socially interact with one another, provide feedback and comment on their drawings, an interaction style that is lost in digital CSTs. In this work, we explore whether collaborating with a robot that also interacts with the creators socially, in a figural co-creation activity benefits children's creativity. We explore whether children emulate the robot's expressed figural creativity through a co-doodling task.

Game Design

To investigate whether children model a social robot's figural creativity, we designed the MagicDraw game, which involves a collaborative drawing interaction on an Android tablet between the child and the robot (Ali et al., 2020). The gameplay requires one player to start a drawing with a stroke, and the other player completing the drawing. After the drawing is complete, the players switch turns. When it is the robot's turn to complete the drawing, we utilize the Sketch-RNN model (Ha and Eck, 2017), which generates drawing strokes to convert a starting stroke into a meaningful illustration (**Figure 3**).

Interaction Scenario

The collaborative figural activity utilized Jibo for similar reasons to the collaborative verbal activity, and its interactions were designed to evoke an autonomous "artistic" peer, collaboratively creating drawings with the child. Even though Jibo does not have appendages, it conveys interest and intent by looking at the tablet while it "draws," and vocalizes relevant phrases, such as "OK, a cat. I think I can make that drawing into a cat." "Here I go!," or "Watch me convert your doodle into a cat." The robot can also ask the child for feedback, e.g. "What do you think about my drawing?," increasing the credulity of the interaction. We verified in the post-test that children perceived Jibo was drawing on the tablet with them. Similar to the Droodle Creativity Game, subjects and Jibo played the figural creativity game by taking turns on an Android tablet.

Experiment Design

Participants

We recruited 78 children in the 5–10 yr age range as a part of the Somerville after-school activities program at the public schools in Somerville, MA. Eleven students in the figural creativity study were excluded due to incomplete data collections or network errors.

Pre-test

All students completed the TTCT assessment as a part of the pretest activity. As with the prior study, the purpose of conducting the TTCT was to drive a quasi-random assignment into groups such that their creativity scores are counterbalanced across the two conditions, described in the following section.

Study Conditions

Participants were divided into two study condition groups: one that interacted with the high creative robot (C+) and one that interacted with the low creative robot (C−). The groups were divided such that the participants in the two groups were balanced in terms of their TTCT scores (C+: 43.33 ± 6.30; C−: 42.91 ± 5.16), age (C+: 7.89 ± 1.91; C−: 7.09 ± 1.96) and gender (C+: F = 14, M = 23; C−: F = 20, M = 21).

The robot in the C+ condition produced more creative drawings as defined by the metrics of the Test of Creative Thinking - Drawing production (TCT-DP) -- a figural creativity test (Urban, 2005). During the robot's turn, we adjusted the drawing model to reduce the randomness in drawing. We kept the speed of drawing to the default speed (60 fps). The robot always drew true to the selected category. This led to higher quality drawings with a better model match to the category that the child selects. The length and number of interactions were controlled for across the two conditions. We validate this hypothesis of these drawings being rated as more creative in the following section.

In the low creative robot condition, the robot system was configured to produce less creative drawings as measured by the TCT-DP figural test parameters. We adjusted the generative model to increase the randomness of the drawing, thereby producing lower quality drawings with a lower model match to the category that the child selects. Further, we adjusted the frame rate of rendering to 30 frames per second to generate the drawings more slowly. We also made the model periodically select an incorrect category to make the drawing not match the selected theme.

Hypotheses

H1: The drawing model produces more creative drawings in the C+ condition than in the C− condition.

H2: Children who played the Magic Draw with the high creative Jibo (C+) will exhibit higher levels of creativity in their own drawings than children that play with the low creative Jibo (C−).

Data Collection and Measures

The MagicDraw application logged the drawings done by the child from all three rounds, the drawings done by the robot in all three rounds, and the time taken for each drawing onto a log file downloaded to the Android tablet. We also used an overhead GoPro camera to take a birds-eye video of the interaction, as well as for recording audio and participants' post-test interviews.

To assess figural creativity from children's drawings in the MagicDraw interaction, we used the Test of Creative Thinking - Drawing Production (Urban, 2005). Three coders blind to the study's hypothesis and the participants' study condition reviewed the drawings and rated them. These scores were then used for calculating the TCT-DP measures of the drawings. Some participants did not make any drawings, and some drawings were not saved due to network errors.

TABLE 3 | MAGICDRAW figural creativity game one-way ANOVA results per condition for each study hypothesis.

Model Condition	Robot Drawing TCT-DP scores (H1)	Children's Drawing TCT-DP scores (H2)
High creative model (C+) (n = 37)	28.91 ± 6.69	42.27 ± 14.30
Low creative model (C–) (n = 41)	20.33 ± 7.86	32.88 ± 9.64
Result (unpaired t-test)	$t(26) = 1.60$, $p = 0.0064$	$t(56) = 1.67$, $p = 0.0023$

One way ANOVA tests revealed that the study condition has a significant effect on participants' figural creativity.

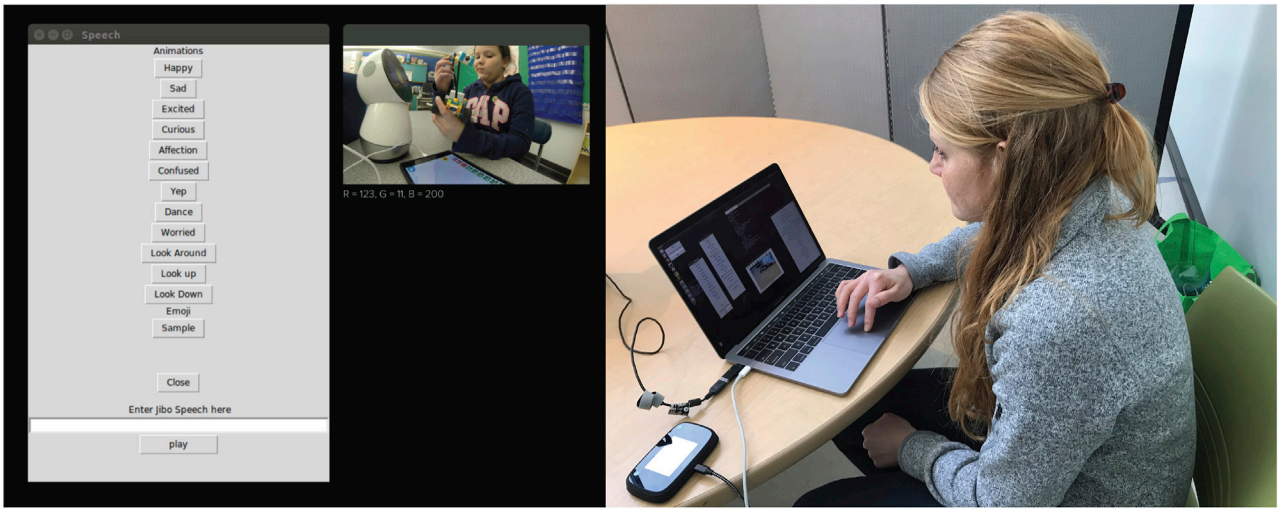


FIGURE 4 | (A) Version 1 of the Jibo control interface for generating Jibo speech commands and emotional expressions. **(B)** A teacher using the remote control interface to control the robot.

Results

We conducted unpaired t-tests comparing the creativity scores of drawings generated by participants in the control condition (C–) and the experimental condition (C+), as measured by the TCT-DP test.

H1: The robot's drawing model produces more creative drawings in the C+ condition than in the C– condition.

To test this hypothesis, we compared the TCT-DP scores of the robot drawings generated by the creative model and by the low creative model. An unpaired t-test showed that the model type had a significant effect ($p < 0.01^{**}$) on the generated drawing's corresponding creativity score (**Table 3**). This dataset was notably smaller than the children's drawing dataset since we did not collect all of the drawings generated by the model. This significant difference helped establish that the creative model was indeed generating drawings that were more creative than the low creative model. Hence, manipulating certain parameters of the model led to a change in the drawing's creativity.

H2: Children who played the Magic Draw with the high creative Jibo (C+) will exhibit higher levels of creativity in their own drawings than children that play with the low creative Jibo (C–).

To test this hypothesis, we compared the TCT-DP scores of all participants in all the conditions. An unpaired t-test test revealed that the study condition had a significant effect on children's figural creativity, and showed a significant difference between the High creative robot (C+) and Low creative robot (C–) ($p < 0.01^{**}$) (**Figure 4**). We could hence validate our second hypothesis.

EXPERIMENT 3: WEDO CONSTRUCTION TASK

Physical construction, or the ability to make new artifacts by combining other artifacts, is a key indicator of children's creativity. Furthermore, unique ways of combining and using the creations also indicates divergent thinking. Physical construction kits facilitate playful learning, open-ended making and creativity, and often creates co-creation space with others (Alimisis, 2013). Robotic toolkits and environments have been successfully leveraged to afford construction by children (Mioduser and Levy, 2010). When children construct with others, their creativity is scaffolded by other children acting as models to emulate, providing ideas, brainstorming, challenging and asking questions. In this experiment, we explore whether this social support can be offered by a social robotic collaborator in a construction activity.

FIGURE 5 | (A) The Jibo control interface. The questions, creative prompts, and positive reinforcement prompts only appeared for the instructors in C+ condition. **(B)** A teacher using the remote control interface to control the robot.

To afford constructional creative expression through making, we designed a third activity in which children and Jibo collaborate building models using the LEGO Education WeDo 2.0 Core Set (2020). The set consists of LEGO bricks and electronics that can be programmed using a visual programming interface on a tablet application with the aim of introducing children to computational thinking and engineering principles in a fun and engaging way.

The interaction involved children making construction projects using the WeDo 2.0 kit in the presence of a social robot, which assumed the role of a tutor and provided scaffolding to the child. In order to get acquainted with the programming interface, children were first guided by the robot to build a rover using LEGO blocks, and to program the rover such that it could detect obstacles and respond to their commands using the WeDo Android tablet application. Children could utilize WeDo 2.0 standard construction kit items, including a Bluetooth enabled controller, LEGO bricks, motors and supporting construction materials and motion sensors. This guided activity was conducted through a step-by-step verbal exchange between the child and the robot and lasted for 6 min, with the robot taking the instructor role (**Figure 5**). The activity introduced the child to sequential commands, condition statements, delays, and loops. Then, children were given 20 min for free play, where they could explore different functions of the WeDo app, add new LEGO blocks, and make their models perform new actions. The idea generation process was guided by both the child and the robot. The role of the robot was to scaffold the child's creative learning through verbal and non-verbal behaviors. Throughout the interaction, children could ask the robot questions and receive dynamic troubleshooting guidance. The robot also engaged in active creativity scaffolding which involved asking the child reflective questions, challenging their ideas and assumptions, and suggesting alternate ideas for creations with the rover. The robot also provided feedback and positive affirmation after children generated new ideas.

While the Droodle Creativity game and MagicDraw game utilized a fully autonomous Jibo interaction, in this activity, the robot was controlled by a human instructor using a dynamic and predictive Graphical User Interface (GUI) on the desktop to provide quality scaffolding at the right time and also to respond to the child appropriately. This GUI controller was iteratively co-designed in tandem with the instructors, with the goal of assisting them in providing creativity scaffolding to the children. This iterative process of designing the GUI is outlined in **Section 3.4.2**. The desktop GUI application communicated with the robot using Robot Operating System (ROS, 2021). Children programmed the WeDo controllers using an Android application on a tablet screen. **Figure 5** illustrates the system components and communications between them.

Interaction Scenario

Introduction

Jibo guides the activity with the child, starting with a self-introduction and then engaging in a short ice-breaker activity. Jibo begins by looking at the child and saying, "Hi. My name is Jibo. What is your name?." When the child responds with their name, Jibo replies with an affectionate expression while saying, "It is so nice to meet you. My favorite activity is to do my favorite dance! What activities do you like?"

Jibo explains the activity to the child: "Today, we will be programming this rover robot to do cool things. [looks down at the tablet] Are you ready to begin?" When the child says "yes," Jibo responds with excitement, "Let's go!"

Child-robot Co-play

Jibo then begins a step-by-step guided activity to help the children learn how to program a rover with the LEGO WeDo kit. This activity ensures that children understand how to use the WeDo construction kit. After this guided activity, children are free to

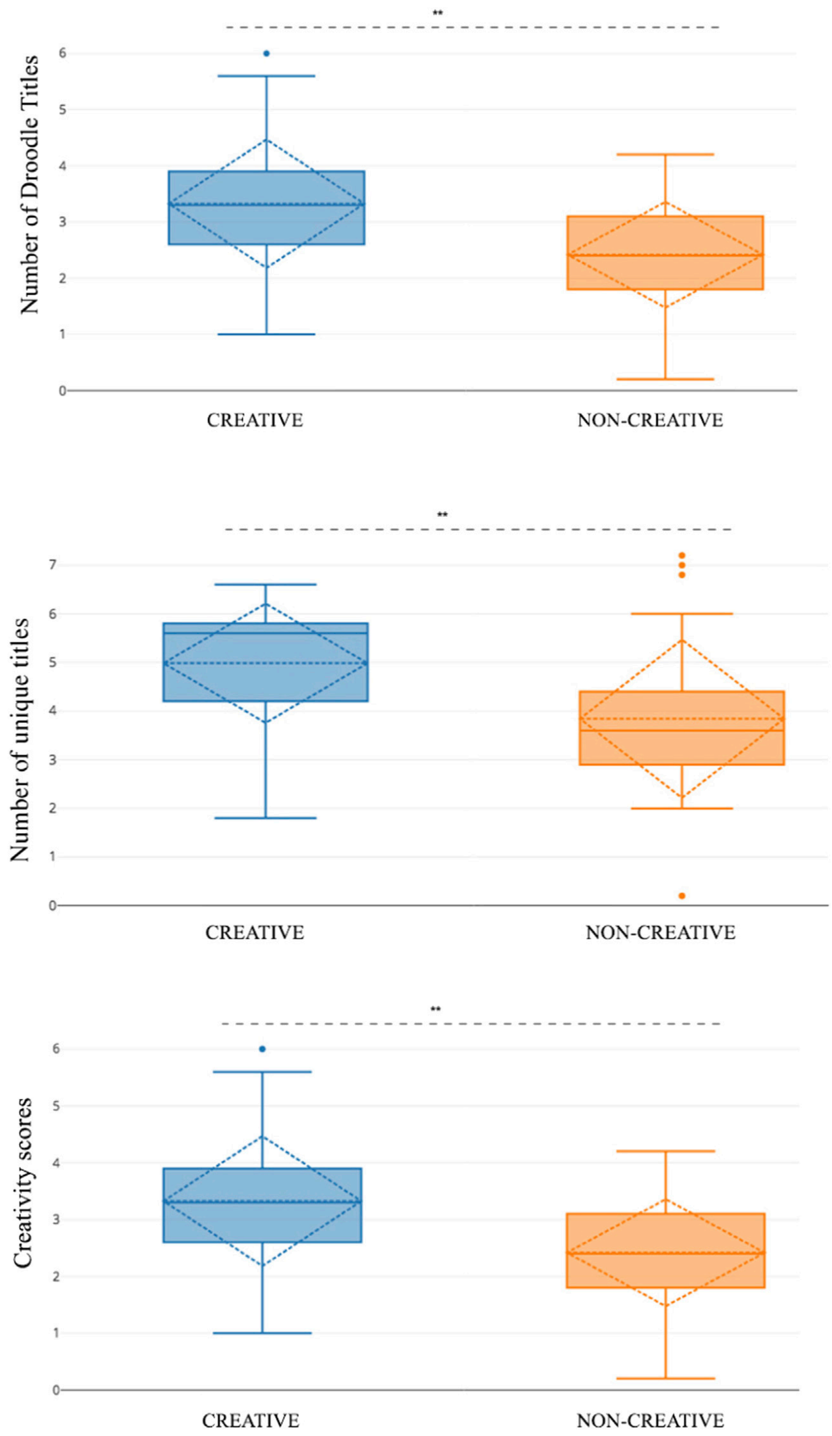


FIGURE 6 | Participants in the C+ condition group generated significantly more Droodle titles, more unique titles and highly creative titles as compared to the C− group.

explore and build new models with the WeDo kit. Jibo provides creativity scaffolding to help children generate novel ideas by using verbal phrases like "Can you think of another way to do this?"

Robot Interactions

Jibo acts as an instructor by: *1*) assisting children in learning how to use the WeDo construction kit, and *2*) scaffolding children's creativity while they construct models. While the tone of

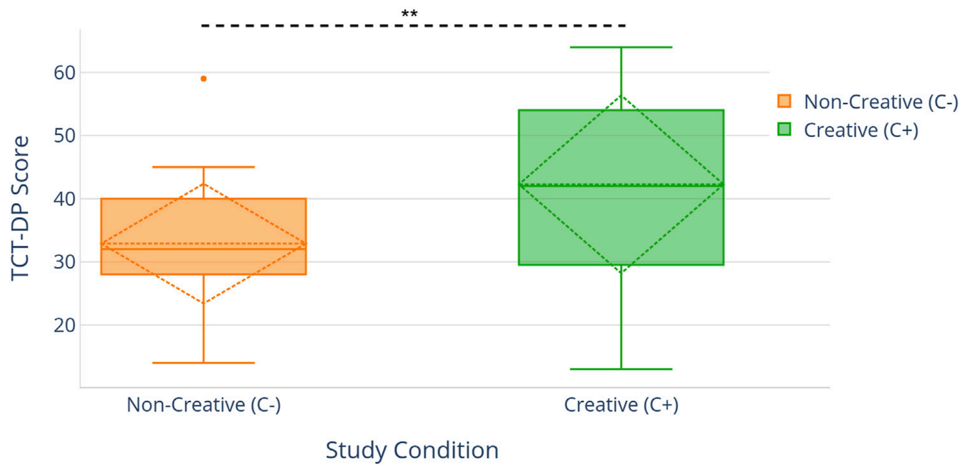


FIGURE 7 | Participants who interacted with the high creative robot scored significantly higher on the TCT-DP test compared to the low creative robot condition (H2).

interaction is collaborative -- for instance, Jibo says, "Today, we will be building a rover together" -- Jibo primarily takes on the role of a tutor that is helping children to create something. Jibo interacts with children through speech prompts and emotional expressions. Jibo's behaviors are remote controlled by a human instructor in a Wizard-of-Oz (WoZ) manner.

Actions for creativity scaffolding are inspired by how human instructors and peers scaffold children and enable them to be more creative by asking reflective questions, generating multiple diverse ideas, challenging assumptions, providing feedback, and appreciating the value of the child's ideas. Creativity and divergent thinking literature elaborates on how asking reflective questions, presenting challenges, and positive reinforcement fosters creativity in children and adults (Kafai et al., 1995; Halverson and Sheridan, 2014). Collaboration with peers and tutors is also beneficial for creative thinking (Rojas-Drummond, et al., 2008). In this activity, instructors use a remote control GUI that has preset suggestions for prompts. We gathered the prompts from collecting and categorizing interaction data from human instructors scaffolding children's creativity using an open-ended WoZ interface. The design of the interface is described in detail in the following section.

WoZ Creativity Scaffolding Interface Design

The goal of designing a robot control interface was to provide instructors with a user-friendly tool to remotely control Jibo while enabling them to provide creativity scaffolding to children. First, three instructors were given a fully flexible desktop interface which contained a text box where they could freely create dialogues and buttons to choose from preset animations for Jibo, while overseeing the interaction using a birds-eye camera view (**Figure 6**). All instructors were trained collectively to understand the task and the WeDo construction kits. The instructors were told that their task was to guide the children to build a simple rover, and then to assist children in thinking creatively about building other WeDo models. Further, all instructors were given a detailed protocol guide for the graphical interface to control Jibo. We evaluated the interface with three instructors and eight participants (6–10 yr old).

An affinity diagramming method was used to categorize all of the prompts that were used by the instructors. And resulted in the following categories:

- Instructions: Construction and programming instructions with the goal of teaching children how to use the WeDo construction kits and build their rover. Instructors tended to use the same language of instruction that was provided to them in the protocol.
- Questions: Reflective questions that the instructors asked children. For example, "Can you tell me why you did that (last action)?," "Can you think of another way to do that (last action)?," or "How will you do that?"
- Creativity prompts: All prompts that were not direct instructions to children but were focused towards helping them come up with creative ideas. These included new ideas and challenges. Prompts included "Can you think of another way to use that block?" or "What else can you make with the same blocks?"
- Feedback: All responses to children's actions. These were mostly positive feedback, such as, "Good job," but also involved encouragement prompts such as "Let's try again."
- Frequently asked questions (FAQ): There were times when children asked the robot questions to help them troubleshoot problems. Some patterns that arose were difficulty connecting the rover to Bluetooth, or not being able to find a part. The responses were first grouped by topics such as "Bluetooth," or "Missing parts" and then all of the instructors' responses to these questions were grouped under FAQ.

We then provided teachers with a more structured GUI for interaction, where frequently used speech prompts were made into buttons in order to reduce time delays, and organized by their categories, as shown in **Figure 7**. Instructors could also input custom speech if needed. We evaluated the interface using a think-aloud evaluation protocol (Ericsson and Simon, 1980), where the instructors spoke about what actions they want to perform, how they use the interface to perform it, and what the interface does not allow them to do. At the end of the interaction, the instructors were asked the following questions:

- What worked in the interface to help you give instructions and scaffold for creativity?
- What did not work in the interface?
- Were there parts of the interface that you did not understand the functionality of?
- What will you change in the interface to better suit your needs?

Based on their interactions with the interface and feedback post interaction, we iterated on the interface design. For instance, an instructor expressed that the interface was too cluttered, and short "clues" to the prompts would be preferable compared to the entire prompt for ease of use. Another instructor expressed how getting a sense of time elapsed on the screen can better prepare them in planning the activity. We stopped the iteration loop when the instructors reported that the interface presets were sufficient for their needs barring some outlier interactions. The final set of creativity scaffolding prompts used in the interface buttons are listed below.

Reflective Questions:

- Can you tell me why you did that?
- What will you be doing next?
- What are you trying to make?
- How are you going to do that?
- What are the materials you would be needing for that?
- Do you have any questions about this?
- Is that the best way to do that?

Creative Prompts (Ideas and Challenges):

- What are some other things you can make the rover do?
- What else can you make the rover do when an object is near? Can you make it have a different output?
- What are some other uses for the [motor/sensor]?
- Let's think of some fun uses of the rover.
- There might be a better way to program that.
- I have an idea!
- Let's try to make an obstacle for the rover's sensor to detect. You can use LEGO blocks to make an obstacle.
- Let's try to make the rover move when you wave your hand in front of the sensor, and stop when you wave your hand again?

Positive Reinforcement:

- That is such a great idea! Good job.
- You think of some really cool use of the robot.
- Well done. That was so creative.
- Great job!
- You are doing so well.
- I would not have thought of that. Good going.

TABLE 4 | Study groups for WeDo Construction task.

Study Groups	*n*	TTCT scores	Gender	Age
High creative (C+)	23	42.16 ± 7.17	F = 11 M = 12	8.3 ± 1.57
Low creative (C–)	20	40.66 ± 6.01	F = 8 M = 12	7.65 ± 1.85

43 participants were divided in balanced groups based on TTCT scores, gender and age.

We conducted 20 playtests, logged instructor's interactions with the GUI, and coded them with: *1*) the previous GUI interaction, and 2) the child's actions that led to the interaction. We then calculated the probability of each GUI interaction following each child action, or preceding GUI interaction. For instance, the child connecting the sensor to the rover body had the highest probability to be followed by the prompt "Instruction: try out the sensor by waving your hand" (Prob. = 0.58). We used these probabilities to predict what the instructor's next interaction with the GUI would be based on the child's actions and the instructor's previous GUI interaction. We developed a dynamic predictive suggestions feature where the interface would prompt the instructor with GUI elements to use when the child performed a certain action. Instructors could choose to use the predicted prompt or create a new one.

Experiment Design

Participants

A total of 43 participants in the 6–10-yr-old age group were recruited for our third study (19 female, 24 male). All students completed the TTCT as a part of the pretest activity. The average age of the participants was 8.11 (S.D. = 1.68). The subjects were recruited as part of the Somerville after-school activities program at the public schools in Somerville, MA. All participants and their guardians signed a consent form to participate and for audio and video data collection. Three adult instructors were also recruited for the study. All instructors were given preliminary training of WeDo construction kits, the programming interface, and a study protocol.

Pre-test

Participants were administered the first part of the verbal and figural module of the TTCT.

Study Conditions

Participants were divided into two study-condition groups: one that interacted with the robot offering creativity scaffolding (C+ condition) and one that interacted with the robot not offering creativity scaffolding (C– condition). The groups were divided such that the participants in the two groups were counterbalanced in terms of their mean and standard deviations of TTCT scores, age and gender (**Table 4**).

In both study conditions, the robot was controlled by a human instructor (blind to the study condition) using a WoZ desktop interface (**Figure 7**). In both conditions, instructors were instructed to start with providing the same set of basic instructions to help the child build and program a rover model that incorporates a sensor and a motor, after which they are left to explore and make their own models. Instructors sent construction commands provided in the WoZ interface after the child completed the previous instruction, as was indicated by the live camera field. Instructors in the creativity scaffolding condition (C+) were instructed to ask reflective questions, challenge the participants, and collaboratively ideate with them. In order to facilitate this scaffolding, they are equipped with the creativity scaffolding interface, which in addition to the construction instructions, also consisted of questions, creative prompts and positive reinforcement prompts (as outlined in **Section 3.4.2**). In contrast, in the C− condition, the robot prompted the child to explore and make new things in the beginning of the free exploration period, and the instructors were instructed to only participate to answer questions beyond that. In the C− condition, the scaffolding interface also lacked the questions, creative prompts and positive reinforcement prompts. The three instructors were paired with an equal number of C+ and C− condition participants to control for experimenter bias.

Hypothesis for the Creativity Scaffolding Study

In order to understand the effect of creativity scaffolding on children's creativity, we hypothesize that participants who interacted with the robot offering creativity scaffolding exhibit higher levels of creativity in the WeDo construction task. We divide our hypothesis in these parts derived from the three ways of assessing creativity behaviors during the task:

H1: Participants who interact with the robot offering creativity scaffolding (C+) generate a greater number of ideas and use cases for the rover than those who interact with the robot without creativity scaffolding (C−).

H2: Participants who interact with the robot offering creativity scaffolding (C+) use a higher number of new programming blocks (excluding the blocks used in the instructions) than those who interact with the robot without creativity scaffolding (C−).

H3: Participants who interact with the robot offering creativity scaffolding (C+) generate more uncommon ideas than those who interact with the robot without creativity scaffolding (C−).

Post-test

We conducted an open-ended descriptive post-test interview with all participants in order to understand how they perceived the creation process.

Q1. Can you describe what you made today?
Q2. How do you think Jibo was helpful to you?
Q3. How do you think Jibo can be of more help?
Q4. Do you think Jibo had any creative ideas?

In order to provide transparency about the robot's abilities, participants were briefed about the WoZ nature of the study and how the robot was being controlled by human instructors.

Data Collection and Measures

All interactions by the instructors on the desktop app were logged on to the computer along with time stamps. All tablet interactions on the WeDo application by the child were logged on to the Android tablets. We used a birds-eye view camera to record the video and audio of the interaction.

Two reviewers watched videos of the interaction and reported the creativity exhibited and the novelty of ideas. The reviewers were blind to the child's study condition as well as the hypotheses, but were familiar with the WeDo construction activity. We used the *fluency* of ideas, *novelty* of ideas, the Unusual Uses task (Silvia, 2011), and divergent thinking as bases for measuring creativity in this task. The following three behaviors were used as metrics of creativity:

1. *Number of use cases for the rover*. We counted the number of unique applications children came up with in the free exploration time as a measure of creativity. For instance, children utilized the toolkit robot's motion sensor and programmed an obstacle course, or used the waving of their hand to display their image. This measure is inspired by the *fluency* and originality of ideas measured (Runco and Acar, 2012), and was calculated by observing the video stream of the interaction.
2. *Number of new programming blocks used*. The instructions teach children how to use some blocks, such as the condition, motor, sensor, start and stop. Additionally, the WeDo programming interface has many different blocks that can be used in different ways, such as the image block, the sound block, other motor blocks, the text block, loops, etc. This measure is inspired by *originality* as a measure of creativity, and was calculated by analyzing the datalog of the tablet interactions.
3. *Commonality*. For each of the new use cases or application ideas of the rover, we determined how uncommon the idea was. We grouped and coded all ideas that were identical or similar, such as "obstacle course" and "lego path." We then looked at the frequency of that idea in the data. For participants with multiple applications of the rover, we took an average of the two frequencies to report commonality. If an idea is uncommon, or deviates from the typical ideas of the group, they count as more creative (Runco and Acar, 2012). The frequencies of each application idea are inversely proportional to creativity. This measure is inspired by *divergent thinking* measures, which look at deviations from the group's trends.

For condition analysis, we calculated numerical values for each of the three metrics by coding the video recording of each interaction. We conducted the Shapiro-Wilk test to check for normal distribution of the data collected, and then conducted an unpaired *t*-test between the study conditions for each of these measures.

TABLE 5 | WeDo construction task results comparing the fluency and originality of idea, and divergent thinking expressed by participants in the three study groups.

SG	Number of ideas for the rover	Number of new blocks	Frequency of ideas
High creative robot (C+) (n = 23)	1.74 ± 1.28	5.96 ± 1.77	0.82 ± 1.03
Low creative robot (C–) (n = 20)	1.05 ± 0.76	4.75 ± 1.65	1.21 ± 1.08
Result	$t(36) = 1.688$, $p = 0.018$	$t(41) = 1.682$, $p = 0.013$	$t(41) = 1.68$, $p = 0.076$

Participants in the High creative robot condition came up with a significantly higher number of ideas, and used a significantly higher number of programming blocks than the Low creative robot (C–) condition.

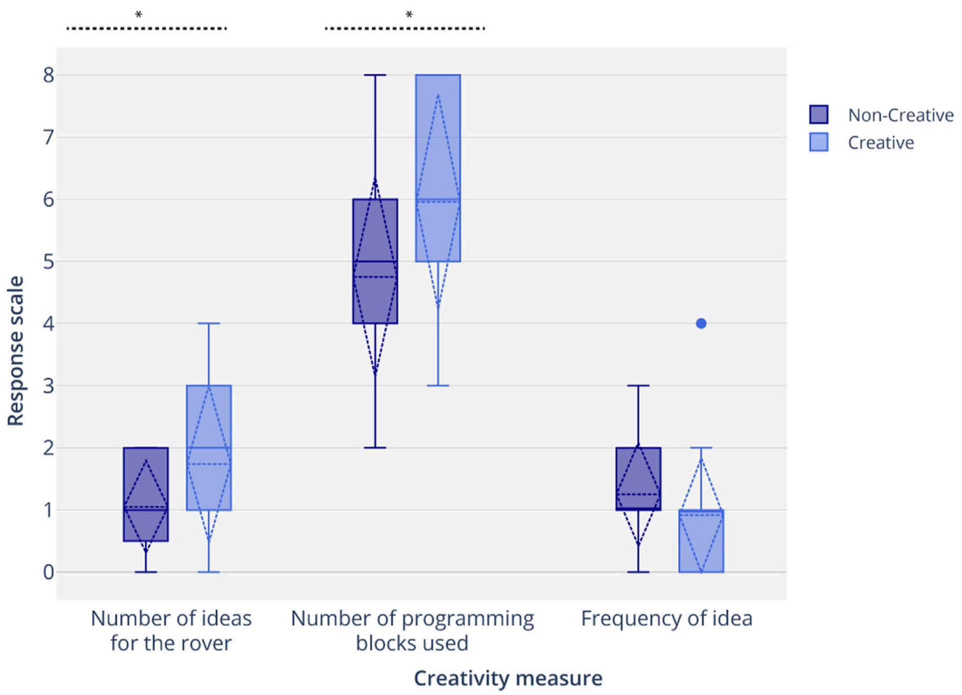


FIGURE 8 | Participants in the High creative robot condition (C+) generated a significantly higher number of ideas than in the Low creative robot (C–) condition (H1). Participants in the C+ condition used a significantly higher number of programming blocks than participants in the C– condition (H2). There was no significant difference in the frequency of ideas generated in the two conditions (H3).

Results

We tested for normality using the Kolmogorov-Smirnov test of normality on all three measures for both groups ({Number of ideas, number of new blocks, frequency of ideas}×{C+, C−}). We found that none of the groups of data differed significantly from that which is normally distributed ($p = 0.22$, $p = 0.57$, $p = 0.55$, $p = 0.36$, $p = 0.26$, $p = 0.053$). Levene's test showed that the variances for the number of ideas were not equal, $F(1,41) = 7.96$, $p = 0.007$, and the variances for number of new blocks, and frequency of ideas were equal, $F(1,41) = 0.312$, $p = 0.602$ and $F(1,41) = 0.277$, $p = 0.579$. For the number of ideas, we conducted the Welch's test assuming unequal variances, which revealed that participants expressed higher *fluency* by generating a significantly greater number of ideas for the rover ($M = 1.42$, $SD = 1.12$, $t(36) = 1.688$, $p = 0.018$). An unpaired parametric *t*-test revealed that participants in the C+ condition expressed significantly higher *originality* by using significantly more number of unique programming blocks ($M = 5.40$, $SD = 1.80$, $t(41) = 1.682$, $p = 0.013$) in the C+ condition as compared to the C− condition (**Table 5**; **Figure 8**). While participants in the C+ condition demonstrated divergent thinking and found unusual uses of the same blocks by expressing less common ideas, or more novel ideas, than the C− condition, this difference was not found to be statistically significant ($M = 1.02$, $SD = 0.96$, $t(41) = 1.68$, $p = 0.076$). *Creativity scaffolding* offered by the robot influenced the number of ideas that children generated and the number of unique programming blocks used. Scaffolding offered by the robot led to more uncommon or atypical ideas by children, but the effect of scaffolding was not significant.

In addition to comparing participants' creativity, we gained an insight into their perception of their creations and the robot's role as a collaborative peer through the post-test questionnaire.

Q1. Can you describe what you made today?

We asked participants to reflect on their creations, what they used and what they learned. One participant said, "*I made him drive, make music and show an image*." Another participant said, "*I made a LEGO robot that when you put your hands to it it'll go back but without you touching it, and then I made a sensor and I made it make a noise*." One participant said they made "*a spaghetti one-eyed snail cricket thing*." These questions helped us unpack not only what their ideas were and which bricks they used, but also what their perceptions of their constructions were. Some participants also spoke about ideas they had but could not construct due to time constraints. One participant said, "*I wanted to make the rover move around and find all the walls, but there was no time*." We also used participants' narratives of what they made to match with the *number of ideas* metric that was reported by the blind reviewer.

Q2. Was Jibo helpful to you? How?

Eighteen participants responded with "yes, he was helpful." Nine participants provided no reasoning. Five participants said "no" (*non-creativity scaffolding* conditions). The most common reasoning response in both study conditions was that Jibo helped them construct the rover by providing instructions. Some participants in the creative condition pointed out how, "*Jibo had cool ideas*" and "*He helped me think of other uses for the sensors*." Multiple participants pointed out how Jibo "*told me when I was doing well, or when he liked my ideas*." Hence, children did notice the positive reinforcement provided by the scaffolding robot. One participant in the C− condition said, "*He kind of was [helpful], but he's super rude. Because half the time I tried to say Jibo, can you help me? He will interrupt me with something else. I tried to say good morning but he didn't reply. it takes him a while to respond like he's not listening*." This highlighted some technical difficulties such as speech delays in implementing this scaffolding model. We also learned that rapport-building utterances, such as greeting utterances in the beginning, can help the children establish common ground with the robot and also help them to get acquainted with speech delays.

Q3. How do you think Jibo can be of more help?

Participants had very valuable feedback about how to make the interaction better. The most common response from both robot conditions was that it would be nicer if the robot displays the blocks to be used on his screen, and that it is difficult to understand which block he is talking about using speech alone. Unlike a human instructor, Jibo cannot point, so visuals would be very helpful. One participant in the C− condition said, "*He could have shown me other things that the rover can do*." One participant in the C+ said, "*He could have told me what a microphone was*." We observed that it is essential to unpack difficult terms that children might not have previously heard.

Q4. Do you think Jibo had any creative ideas?

79% of participants in the C+ condition, and 35% of participants in the C− condition, responded with "*yes*." Hence, participants successfully perceived the expressed creativity of the robot. Some participants went on to explain why they thought that Jibo was creative. One participant said, "*Yes, he told me to make the [rover] move and can put more than one thing on the screen*." Another participant said, "*He had cool ideas like playing music. He played fun games with me and he had great ideas and he knows that he's smart*." One participant also said, "*Jibo thought that I had cool ideas, and that made me happy*", and another one said, "*Yes, he told me I can make what I want and told me my idea was great*." Among participants that responded with "*No*" or "*Maybe*" there was typically no reasoning. One child in the C− condition said, "*Jibo knew what to do but he was not really creative*."

DISCUSSION

In this work, we demonstrate how a social robotic peer's socio-behavioral patterns can influence creativity in children in the 5–10 yr old age group. Specifically, we highlight two robot interaction patterns: *1*) *creativity demonstration*, where the robot itself demonstrates artificial creativity, and *2*) *creativity scaffolding*, where the Jibo robot supports and encourages the child's creative thinking by asking reflective questions, providing challenges and positive reinforcements. We designed three game-based interactions that afforded different types of creativity. The robot's interaction patterns were inspired by children's creativity-eliciting social interactions with their peers and tutors. These interactions serve as playful ways of measuring creativity, as well as supporting children's creative expression. In order to assess the efficacy of these interaction patterns, we conducted an accompanying investigative study for each of our game-based interactions: the Droodle Creativity game that affords children's *verbal* creativity, the MagicDraw game that affords children's *figural* creativity, and the WeDo construction task that affords children's *constructional* creativity.

In the Droodle Creativity game, children emulated artificial verbal creativity exhibited by Jibo during gameplay. Participants who interacted with the high creative robot expressed more ideas, more diverse ideas, and highly creative ideas in the Droodle Creativity game, as compared to participants that interacted with the low creative robot. Similarly, in the Magic Draw Game, children emulated artificial figural creativity exhibited by Jibo in a co-drawing task, which led to their drawings being measured as more creative by the TCT-DP test for figural creativity. Participants also perceived the high creative robot as highly creative. Through these two studies, we verify our first hypothesis: that children adeptly socially emulate the creative behavior of peer-like robot playmates, and this in turn fosters children's own creative behavior. Importantly, this is a sufficiently robust enough finding that we could replicate it for two different kinds of creative expression: verbal and figural creativity.

In the second investigation, we demonstrated how a robot offering creativity scaffolding in the form of asking reflective questions, challenging the participants, and providing positive reinforcement had a positive effect on children's creativity. Participants engaged in an open-ended activity involving constructing and programming a rover using the WeDo construction kit. Creativity was measured by the ideas that children came up with for the rover, the number of different tools (programming bricks) they used, and how unique their ideas were. Children interacting with the robot that offered creativity scaffolding scored significantly higher on the number of different ideas and different programming bricks used. They also scored higher in the uniqueness of their ideas; however, that difference was not statistically significant. Hence, we could establish evidence that children can learn creativity from a social agent by emulating the agent's creative behaviors and by the agent scaffolding their creative learning, which informs the design of pedagogical embodied tools to foster creativity.

The WeDo construction task utilized a predictive scaffolding model. This paves the way toward the development of fully autonomous robot scaffolding systems for tailoring personalized learning to different students and contexts. Over time and over several playtests, the model is able to reinforce itself depending on whether the instructor accepts or rejects its suggestions, eventually leading to minimal error rates. Replicating a human instructors' scaffolding into an artificial agent can be beneficial for personalized assistance when the teacher is not present or when there are many students per teacher. This scaffolding paradigm could be used in the context of other activities.

It is important to be wary of the shortcomings of such a suggestion based system for instructors. While these recommendations make it easy for instructors to provide help and scaffold children's creative learning, they also inhibit the instructor's original thought and manner of scaffolding, which holds high value. A more autonomous model would be built from data collected from multiple instructors with a diverse set of backgrounds and expertise, all of whom could instruct several students from a diverse set of backgrounds. Further, the model should be able to adapt based on an instructor's usage while allowing enough space for original thought. While the current interface does allow instructors to reject the model's recommendation and instead generate new Jibo utterances, it can still influence the instructors' decision making process, and cause them to conform to commonly used instructions, which may be counterintuitive to promoting creativity. In our work we suggest two interaction patterns of social robots that effectively foster creativity in young children: creativity demonstration and creativity scaffolding using questions, questions and ideations. We position social robots as CSTs in collaborative activities with children that can leverage the benefits of having a co-located peer to create and socially interact with, and a digital CST that adapts to the person's creation style. We also add to the literature of HRI suggesting that children emulate a robot's learning behavior and extend it to verbal, figural and constructional creativity. Through our observations from playtests, iterative game design, study results and post-test questionnaires, we formulated evidence-based design guidelines as well as additional recommendations from researchers' reflections for designing creativity supporting social agents that we have outlined in the next section. These recommendations could benefit pedagogical researchers, educators and HRI and HCI practitioners designing social agents for stimulating creativity, especially in children.

Design Recommendations

Evidence-based Design Guidelines

In this section, we suggest the following interaction design recommendations for social agents to support children's creativity based on our empirical findings:

1. The social agents should demonstrate the creativity behaviors that the designers aim to foster. We observed that children can learn verbal and figural creativity by emulating a social robot's creative behaviors across all three tasks. Hence, while designing social robots as pedagogical tools, we must ensure that they express the desired creative behavior that researchers aim to foster in children. The expression of creativity is context-dependent (such as generating creative drawings) and can be supported by social behaviors such as reflecting on the artifact generated, or making the creative process transparent through dialogue.
2. Make use of reflective questioning and challenges. Asking reflective questions about the children's actions aids their metacognition and creative thinking. Through the scaffolding study, we learned that instructors who were controlling the robot remotely chose to use many reflective questions as robot speech prompts to provide creativity scaffolding for children. Providing children with optimal levels of difficulty encouraged them to solve problems creatively. In the WeDo construction activity, the scaffolding robot provided challenges to the child, such as, "Can you think of other uses of the same sensor?" or "Do you think that's the best way to do it?" These questions were followed by students exploring creative uses of the objects beyond the first use that they imagined which encouraged flexibility. For instance, one child utilized the motion sensor's ability to detect motion and turn to create an obstacle path for her rover. We observed that children who interacted with the creativity scaffolding robot, which asked reflective questions, exhibited higher levels of creativity in the task.
3. The agent should generate unique and frequent ideas during the interaction. In addition to asking reflective questions and posing challenges, the robot could also demonstrate new idea generation. During the WeDo task, when the robot suggested, "You can use the picture icon to display images when the robot senses an object," the child subsequently used the feature to display the preset images. The child then uploaded their own image and made the rover move towards them. When the rover detected the child as the obstacle, it displayed the child's photograph. Children not only incorporated the robot's ideas, but also built upon them. In the Droodle Creativity game, we observed that the robot's idea generation behaviors were emulated by the child, both in terms of the fluency and originality of ideas generated.
4. Provide positive reinforcement to children when they create. Children in all three tasks commented how the robot said "Good job!" or other similar positive comments after they completed the task. Positive reinforcement after creative behaviors has a strong influence on children. Children often form relationships with social robots, which lead to increased learning gains (Westlund et al., 2018) and getting

positive validation upon exhibition of creativity encourages them to be more creative.

Additional Recommendations From Researcher's Reflections

We incorporated several design principles from background research in CSTs and iteratively designing our child-robot interactions. While we found these design decisions to be beneficial for children's creativity and had a positive combined effect, we did not analyze the effectiveness of each of these recommendations, and future work is required to disambiguate individual effectiveness. We present these as additional design recommendations for designing creativity scaffolding interventions for children:

1. Co-design interactions with instructors. Co-designing scaffolding robot interactions with instructors helped us personalize interactions for the students, and incorporate instructors' teaching experience into the robot's behaviors. Instructors were particularly helpful in designing the interaction GUI; while the pre-set GUI speech prompts were grounded in historic scaffolding commands used by other instructors, we found that teachers often needed to personalize interactions during unique occurrences. Having the option of typing new commands in the GUI text-box afforded instructors that flexibility. This hybrid interface equipped teachers with historically useful commands, reducing their cognitive load during the task, and gave them the ability to create their own.
2. Scaffolding must be grounded in tasks and materials. In the creativity scaffolding task, we started with providing teachers with generic scaffolding prompts such as "What else can you make?" or "What is another way to do that?" However, teachers provided us with feedback that the scaffolding prompts needed to be specific to the task (construction) and materials (blocks). For instance, they used the prompt, "How else can you use that sensor to make the rover move?" This grounding in the collaborative task portrayed the robot as a context-aware scaffolding agent. Hence, while designing scaffolding interactions such as challenges or reflective questions, we found it beneficial to ground the interactions in the task's context instead of generic interactions. However, this reduces the scalability of these interactions across tasks.
3. Agents must scaffold, but not impose. Scaffolding through social interactions can be powerful but has the potential to inhibit creativity. Interactions such as idea generation in collaborative tasks must be designed such that the agent does not impose their ideas on the child, nor intrude upon the children's creative space. For instance, in the WeDo scaffolding tasks, we observed that teachers who controlled the robot suggested ideas related to the child's working idea, and while the child worked on an idea, they did not interfere. This delicate scaffolding can be a challenge to execute in fully autonomous interactions. Providing scaffolding only when the child asks, or when the child is stuck, could be a beneficial approach. Further, care should be taken to not interrupt children's creative process; in the MagicDraw interaction, where players had a fixed time to draw for each turn, one participant reported displeasure for the robot interrupting their drawing.
4. Design game-based interactions with peer-like social agents. Designing game-based child-robot interactions enabled us to position the robot as a collaborative playful peer. This made the interaction fun for children as reported in the post-test, and children were engaged throughout the games. We designed game tasks with no fail state in order to provide an outlet for unconstrained creative thinking and encourage divergent thinking. Since assessment is shown to hinder creativity, we refrained from providing any assessment during the interaction. Game-like interactions made the tasks engaging for young children and allowed for a safe space for failure.
5. Center task around creation of artifacts. In order to maximize space for creative thinking, we designed the tasks around the creation of artifacts rather than the completion of a specific deliverable. In accordance with literature showing evaluation's negative effect on creativity, these artifacts were designed to have no set of "correct" answers, supporting creative exploration without an end goal. To ground robotic scaffolding in the context of the task, we provided a limited set of materials that can be used creatively to produce an unlimited number of artifacts. For instance, in the WeDo construction task, one student wanted to create two rovers but was limited to one controller. They wired the sensor of the second rover from the first rover's controller and called it a "parasite" rover.
6. Leverage collaboration as an agent behavior and game mechanic. Since collaboration has a positive influence on creativity, we must ensure that the child-agent interactions are collaborative in nature and the robot acts as a collaborative peer instead of a competitive one, which hinders creativity in children. The collaborative nature of the interaction was made explicit in robot speech, such as, "Today we will program a robot together." Within our tasks, we framed the social robot agent as a peer helping the child do their best creative work, and the majority of children perceived Jibo as a collaborator rather than a competitor. Collaboration is among the most prominent social factors that positively influence creativity. Careful consideration must be given to interactions with the agent in particular, in order to ensure that children see it as a partner rather than as a competitor, which can hinder their creativity. Introduction of the robot and the task can be leveraged to position the robot as a collaborative peer.

CONCLUSION

In this work, we posit social robots as CSTs for children in collaborative tasks. We studied the effects of an autonomous social robot's verbal and nonverbal interactions on children's creativity as measured by three collaborative game-based child-robot interactions. We observed that both creativity demonstration and creativity scaffolding offered by the social robot had a positive effect on children's creativity in verbal, figural and constructional creativity tasks. This work contributes to the design of game-based child-robot interactions that afford creativity, provides evidence for the efficacy of these interactions and provides guidelines for designing social embodied agents to foster creativity in young children. These findings are valuable to game designers creating game-based interactions to foster children's creativity, as well as HRI and HCI practitioners leveraging social agents as CSTs.

Since robots are already being used in classroom settings as learning peers and personalized tutors, it is imperative to think about how their behaviors can influence children's learning behaviors, such as creative thinking. While social robots are not the only way to provide creativity support through behavioral modeling, they certainly are a compelling way given their social nature. Effort must go into designing the agents' behavior such that they exhibit creativity and scaffold the child's creativity as a peer or a tutor. Embodied AI agents have the potential to use generative modeling techniques to express different forms of creativity through generating media such as drawing, poetry, art styles, patterns, physical body movements, etc. They are also socially emotive and can express the social interactions that accompany creativity such as reflection, inquisitiveness and positive affect. This work opens up opportunities to explore how these different forms of artificial creativity can be embedded into tools that children use and interact with, and help them be more creative.

Recent works have demonstrated how robots can help creativity, as a co-present partner and through social interactions. In this work, we suggest two interaction patterns of social robots that we observed to effectively foster creativity in young children: creativity demonstration and creativity scaffolding using questions, questions and ideations. We add to the literature of HRI suggesting that children emulate robots' learning behaviors, and that this phenomenon extends to creativity. We also contribute to the field of Creativity Support Tools by positioning social peer-like robots as a creativity support peer in collaborative activities. This contribution is not only valuable for HRI practitioners, but also other interactive AI agents, such as conversational agents. Creativity supporting social robots combine the benefits of having a co-located peer to collaboratively create with and socially interact with, and a digital CST that adapts to the user's creation style. Leveraging generative AI models now allow for expressing creativity in several modalities, which robots can successively leverage. Further, we elaborate the design of a scaffolding mechanism by learning from human scaffolders through one construction activity. This approach can be generalized to other activities.

While introducing an extrinsic factor in the form of a social robot, we must ensure that it does not come across as an evaluator; classroom research has demonstrated how extrinsic factors such as evaluation, competition and unrealistic expectations can potentially inhibit creativity, instead of fostering it (Torrance, 1967). In the scaffolding GUI design, we provide instructors with a predictive interface that helps them scaffold the child for creative learning; however, this suggestion model also limits creativity and personalization of teaching style from the instructors. To tackle this issue, we must also aim to build personalized scaffolding models that take input from every teacher and personalize over time in both the content and style of learning.

In our work, we chose a wide age range (5–10 yr) and we did not analyze differences across age ranges. This is a limitation of the current work, and future analysis needs to be run with narrower age ranges to determine the efficacy of the intervention on different age groups. Another limitation of this work is that these robot interactions lead to an increase in creativity within the narrow constructs of these tasks, and may not scale to every creative task, or to students' life outside of these tasks. There are also countless ways of expressing creativity such as poetry, storytelling, painting, music, etc., that are not explored in these tasks. This work also defines a limited scope of creativity in terms of fluency, novelty and value of ideation. Creativity encompasses a much wider array of behaviors (such as divergent thinking) that can be explored using other interactions. Furthermore, while all these interactions currently focus on one-on-one child-robot interaction, we must strive towards designing interactions that involve multiple children because collaboration with peers forms a major part of creative learning.

Finally, while this work evaluates the role of the robot's creativity fostering behaviors, it does not evaluate the benefits of embodiment over other non-embodied agents such as computers or voice agents. Previous research has found that adults did not show significant gains in creativity merely in the presence of a social robot (Alves-Oliveira et al., 2019). In future work, we aim to study the combined effect of embodiment and creativity scaffolding behaviors by running a 2 × 2 study ({embodied, non-embodied}×{scaffolding, non-scaffolding}) (Devasia et al., 2020). In order to evaluate whether robots are really that social, future work is required to assess the creativity effects of human peers vs robots.

While this work makes use of Jibo as the social agent, these interactions are learning tools that can foster creativity in classrooms and homes independent of Jibo. These games also serve as game-based creativity assessment measures.

We designed a creativity scaffolding paradigm for the WeDo construction task. This model currently supports a semi-autonomous scaffolding system, where a human controls the robot using a remote control desktop program. In the current version of the robot control interface, which lets instructors control the robot remotely, we can incorporate ASR to use the instructors' speech to control the robot's speech. Moreover, collecting more data about how instructors use the program can help us build a fully autonomous model of scaffolding. While this approach can be used to design interactions for an autonomous or semi-autonomous system, the timing of the interactions and during-study improvisation are reliant on children's actions in the task and interactions with the robot. Current status of natural language understanding and computer vision limit a complex understanding on the scene, and hence a similar fidelity of robotic interactions are challenging to currently implement in a fully autonomous system. However, as we demonstrated, a semi autonomous system (where the system suggests interactions and the human decides the timing of the prompts) is feasible.

Another limitation of this work is that all activities are self-contained and involve single interactions. In future work, it would be valuable to evaluate creativity transfer from one activity to another, and even in the absence of the robot in the long-term. Design recommendations from this work can be incorporated in several creativity support tools such as computer games, voice agents, tablet apps, embodied tools, space design, etc. Finally, advances in generative modeling techniques enable us to create child-robot interactions supporting multiple modalities of autonomous creative expression.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Massachusetts Institute of Technology Institute Review Board (IRB). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SA designed the child-robot interactions, conducted the user study, evaluated the data and wrote about the findings. ND assisted in conducting the study and writing about the findings. HP and CB were advisors for interaction and study design, and evaluation of the data collected.

FUNDING

This work is funded by the MIT Media Lab Consortia fund.

ACKNOWLEDGMENTS

We would like to thank the students and teachers of Somerville Public Schools for their participation in this study. We would also like to thank Ellie Hoban, Randi Williams, Tyler Moroso and Marwa AlAlawi for helping us conduct studies and collect data.

REFERENCES

Adamson, T., Ghose, D., Yasuda, S. C., Shepard, L. J. S., Lewkowicz, M. A., Duan, J., et al. (2021). "Why We Should Build Robots that Both Teach and Learn," in Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21), Boulder, CO, March 8–11, 2021. Available at: https://scazlab.yale.edu/sites/default/files/files/hrifp1028-adamsonA.pdf. doi:10.1145/3434073.3444647

Ali, S. A. (2019). *Designing Child Robot Interaction for Facilitating Creative Learning*. Cambridge, MA: Massachusetts Institute of Technology.

Ali, S., Moroso, T., and Breazeal, C. (2019). "Can Children Learn Creativity from a Social Robot?," in Proceedings of the 2019 on Creativity and Cognition, (San Diego, CA: IEEE), 359–368.

Ali, S., Park, H. W., and Breazeal, C. (2021). A Social Robot's Influence on Children's Figural Creativity during Gameplay. *Int. J. Child-Computer Interaction* 28 (100234), 100234. doi:10.1016/j.ijcci.2020.100234

Ali, S., Park, H. W., and Breazeal, C. (2020). "Can Children Emulate a Robotic Non-player Character's Figural Creativity?" in Proceedings of the Annual Symposium on Computer-Human Interaction in Play (IEEE), 499–509. doi:10.1145/3410404.3414251

Alimisis, D. (2013). Educational Robotics: Open Questions and New Challenges. *Themes Sci. Tech. Edu.* 6 (1), 63–71.

Alves-Oliveira, P., Arriaga, P., Cronin, M. A., and Paiva, A. (2020). "Creativity Encounters Between Children and Robots," in Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20), Cambridge, UK, March 23–26, 2020. (IEEE), 379–388. doi:10.1145/3319502.3374817

Alves-Oliveira, P., Tulli, S., Wilken, P., Merhej, R., and Gandum, J. (2019). Sparking Creativity with Robots: A Design Perspective. OSF [Preprint]. Available at: https://osf.io/preprints/za5h8/ (Accessed October 20, 2019).

Amabile, T. M., and Gryskiewicz, N. D. (1989). The Creative Environment Scales: Work Environment Inventory. *Creativity Res. J.* 2 (4), 231–253. doi:10.1080/10400418909534321

Anki (1999). Cozmo. Available at: https://www.anki.com/en-us/cozmo (Accessed February 21, 2021).

Baas, M., De Dreu, C. K. W., and Nijstad, B. A. (2008). A Meta-Analysis of 25 Years of Mood-Creativity Research: Hedonic Tone, Activation, or Regulatory Focus? *Psychol. Bull.* 134 (6), 779–806. doi:10.1037/a0012815

Baggerly, J. N. (1999). Adjustment of Kindergarten Children through Play Sessions Facilitated by Fifth Grade Students Trained in Child-Centered Play Therapy Procedures and Skills. Dissertation (Denton, TX).

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social Robots for Education: A Review. *Sci. Robot.* 3 (21), eaat5954. doi:10.1126/scirobotics.aat5954

Berretta, S., and Privette, G. (1990). Influence of Play on Creative Thinking. *Percept Mot. Skills* 71 (2), 659–666. doi:10.2466/pms.1990.71.2.659

Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms*. London and New York: Routledge.

Bornstein, M. H., and Bruner, J. S. (2014). *Interaction in Human Development*. London and New York: Psychology Press.

Bowman, N. D., Kowert, R., and Ferguson, C. J. (2015). "The Impact of Video Game Play on Human (And Orc) Creativity," in *Video Games and Creativity*. Editors G. Green and J. C. Kaufman (Academic Press), 41–56. doi:10.1016/b978-0-12-801462-2.00002-3

Carterette, E. C., Friedman, M., Miller, J. L., and Eimas, P. D. (1994). *Handbook of Perception and Cognition*. New York, NY: Academic Press.

Chen, H., Park, H. W., and Breazeal, C. (2020). Teaching and Learning with Children: Impact of Reciprocal Peer Learning with a Social Robot on Children's Learning and Emotive Engagement. *Comput. Edu.* 150, 103836. doi:10.1016/j.compedu.2020.103836

Claxton, A. F., Pannells, T. C., and Rhoads, P. A. (2005). Developmental trends in the creativity of school-age children. *Creativity Res. J.* 17 (4), 327–335.

Dansky, J. L. (1980). Make-Believe: A Mediator of the Relationship between Play and Associative Fluency. *Child. Dev.* 51, 576–579. doi:10.2307/1129296

Devasia, N., Ali, S., and Breazeal, C. (2020). "Escape! Bot: Child-Robot Interaction to Promote Creative Expression During Gameplay," in Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20), 219–223. doi:10.1145/3383668.3419895

Duncan, S. C. (2011). Minecraft, beyond construction and survival. Well Played: a journal on video games, value and meaning. 1:1. Pittsburgh, PA.

Ericsson, K. A., and Simon, H. A. (1980). Verbal Reports as Data. *Psychol. Rev.* 87 (3), 215–251. doi:10.1037/0033-295x.87.3.215

Frich, J., MacDonald Vermeulen, L., Remy, C., Biskjaer, M. M., and Dalsgaard, P. (2019). "Mapping the Landscape of Creativity Support Tools in HCI," in Proceedings of CHI Conference on Human Factors in Computing Systems (CHI '19), Glasgow, UK, May 4–9, 2019. (New York, NY: ACM), 1–18. doi:10.1145/3290605.3300619

Garaigordobil, M. (2006). Intervention in Creativity with Children Aged 10 and 11 years: Impact of a Play Program on Verbal and Graphic-Figural Creativity. *Creativity Res. J.* 18 (3), 329–345. doi:10.1207/s15326934crj1803_8

Gardner, H., and Art, M. (1982). *Brain: A Cognitive Approach to Creativity*. New York, NY: Basic Books.

Ge, S., Goswami, V., Zitnick, C. L., and Parikh, D. (2020). Creative Sketch Generation. Vienna, Austria.

Google (2020). Google. Available at: https://cloud.google.com/speech-to-text/.

Gordon, G., Breazeal, C., and Engel, S. (2015). "Can Children Catch Curiosity from a Social Robot?" in HRI'15: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (Portland, OR: IEEE), 91–98.

Guilford, J. (1950). *Creativity*. American Psychology. 5 (9), 444–454. https://doi.org/10.1037/h0063487

Guilford, J. P. (1957). Creative Abilities in the Arts. *Psychol. Rev.* 64 (2), 110–118. doi:10.1037/h0048280

Ha, D., and Eck, D. (2017). A Neural Representation of Sketch Drawings. arXiv [Preprint]. Available at: http://arxiv.org/abs/1704.03477.

Hallgren, K. A. (2012). Computing Inter-rater Reliability for Observational Data: An Overview and Tutorial. *Tqmp* 8 (1), 23–34. doi:10.20982/tqmp.08.1.p023

Halverson, E. R., and Sheridan, K. (2014). The Maker Movement in Education. *Harv. Educ. Rev.* 84 (4), 495–504. doi:10.17763/haer.84.4.34j1g68140382063

Harel, I. E., and Papert, S. E. (1991). Constructionism. Available at: https://psycnet.apa.org/fulltext/1991-99006-000.pdf.

Henriksen, T. D. (2006). "Games and Creativity Learning," in *Role, Play, Art: Collected Experiences of Role-Playing*. Editors T. Fritzon and T. Wrigstad (Stockholm, Sweden: Föreningen Knutpunkt), 3–15.

Howard-Jones, P., Taylor, J., and Sutton, L. (2002). The Effect of Play on the Creativity of Young Children during Subsequent Activity. *Early Child. Dev. Care* 172 (4), 323–328. doi:10.1080/03004430212722

Jibo (2015). *jibo.com*.

Kafai, Y. B. (1995). *Minds in Play: Computer Game Design as a Context for Children's Learning*. Hillsdale, NJUnited States: L. Erlbaum Associates Inc..

Kahn, P. H., Jr, Friedman, B., Severson, R. L., and Feldman, E. N. (2005). *Creativity Tasks and Coding System-Used in the Plasma Display Window Study*. Seattle, WA: University of Washington. Available at: https://www.researchgate.net/profile/Rachel_Severson/publication/33513295_Creativity_Tasks_and_Coding_System_–_Used_in_the_Plasma_Display_Window_Study/links/560d746608ae96742010cd80.pdf.

Kahn, P. H., Kanda, T., Ishiguro, H., Gill, B. T., Shen, S., Ruckert, J. H., et al. (2016). "Human Creativity Can Be Facilitated through Interacting with a Social Robot," in 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, New Zealand, March 7–10, 2016. (IEEE), 173–180. doi:10.1109/HRI.2016.7451749

Kaufman, J. C., and Sternberg, R. J. (2010). *The Cambridge Handbook of Creativity*. Cambridge, UK: Cambridge University Press.

Kim, H. J., Park, J. H., Yoo, S., and Kim, H. (2016). Fostering Creativity in Tablet-Based Interactive Classrooms. *J. Educ. Tech. Soc.* 19 (3), 207–220.

LEGO Education WeDo 2.0 Core Set (2020). LEGO. Available at: https://education.lego.com/en-us/products/lego-education-wedo-2-0-core-set/45300#wedo-20. (Accessed February 21, 2021).

Ligthart, M. E. U., Neerincx, M. A., and Koen, V. (2020). "Design Patterns for an Interactive Storytelling Robot to Support Children's Engagement and Agency," in Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (Cambridge, United Kingdom) (HRI '20) (New York, NY: Association for Computing Machinery), 409–418. doi:10.1145/3319502.3374826

Louis, M., and Peter, M. (2015). "Face the Music and Glance: How Nonverbal Behaviour Aids Human Robot Relationships Basedin Music," in Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (Portland, Oregon, USA) (HRI '15) (New York, NY: Association for Computing Machinery), 237–244.

MacDonald, R., Miell, D., and Morgan, L. (2000). Social Processes and Creative Collaboration in Children. *Eur. J. Psychol. Educ.* 15 (4), 405–415. doi:10.1007/bf03172984

Mellou, E. (1995). Review of the Relationship between Dramatic Play and Creativity in Young Children. *Early Child. Dev. Care* 112 (1), 85–107. doi:10.1080/0300443951120108

Miell, D., and MacDonald, R. (2000). Children's Creative Collaborations: The Importance of Friendship when Working Together on a Musical Composition. *Soc. Dev.* 9 (3), 348–369. doi:10.1111/1467-9507.00130

Mioduser, D., and Levy, S. T. (2010). Making Sense by Building Sense: Kindergarten Children's Construction and Understanding of Adaptive Robot Behaviors. *Int. J. Comput. Math. Learn.* 15 (2), 99–127. doi:10.1007/s10758-010-9163-9

Mumford, M. D. (2001). Something Old, Something New: Revisiting Guilford's Conception of Creative Problem Solving. *Creativity Res. J.* 13 (3-4), 267–276. doi:10.4324/9781410608604-4

Park, H. W., Rosenberg-Kima, R., Rosenberg, M., Gordon, G., and Breazeal, C. (2017). "Growing Growth Mindset with a Social Robot Peer," in Proceedings of the ACM SIGCHI (Vienna, Austria: ACM Conference on Human-Robot Interaction), 137–145. doi:10.1145/2909824.3020213

Plucker, J. A., Beghetto, R. A., and Dow, G. T. (2004). Why Isn't Creativity More Important to Educational Psychologists? Potentials, Pitfalls, and Future Directions in Creativity Research. *Educ. Psychol.* 39 (2), 83–96. doi:10.1207/s15326985ep3902_1

Price, R. (1982). *Droodles*. Los Angeles, CA: Pricd.

Price, R., and Lovka, B. (2000). *Droodles: The Classic Collection*. Los Angeles, CA: Tallfellow Press.

Rababah, L. M., Alshehab, M. H., and Melhem, N. Z. B. (2017). "Exploring the Factors that Hinder Jordanian Students in Developing Creativity in EFL Writing," in International Conference on Studies in Arts, Humanities and Social Sciences (SAHSS-2017), Bali, Indonesia, January 31–February 1, 2017. (IEEE), 7–11. doi:10.17758/eirai.f0117427

Rojas-Drummond, S. M., Albarrán, C. D., and Littleton, K. S. (2008). Collaboration, Creativity and the Co-construction of Oral and Written Texts. *Thinking skills and creativity* 3 (3), 177–191. doi:10.1016/j.tsc.2008.09.008

Rond, J., Sanchez, A., Berger, J., and Knight, H. (2019). "Improve with Robots: Creativity, Inspiration, Co-performance," in 2019 28th IEEE InternationalConference on Robot and Human Interactive Communication (RO-MAN) (New Delhi, India: IEEE), 1–8.

Root-Bernstein, R., and Root-Bernstein, M. (2017). "People, Passions, Problems: The Role of Creative Exemplars in Teaching for Creativity," in *Creative Contradictions in Education: Cross Disciplinary Paradoxes and Perspectives*. Editors R. A. Beghetto and B. Sriraman (Springer International Publishing), 143–164. doi:10.1007/978-3-319-21924-0_9

ROS (2021). *Robot Operating System*. Available at: ros.org. (Accessed February 21, 2021).

Runco, M. A., and Acar, S. (2012). Divergent Thinking as an Indicator of Creative Potential. *Creativity Res. J.* 24 (1), 66–75. doi:10.1080/10400419.2012.652929

Russ, S. W. (2003). Play and Creativity: Developmental Issues. *Scand. J. Educ. Res.* 47 (3), 291–303. doi:10.1080/00313830308594

Scratch Team, M. (2017). *Scratch Coding Cards: Creative Coding Activities for Kids*. San Francisco, CA: No Starch Press.

Shneiderman, B. (2002). Creativity Support Tools. *Commun. ACM* 45 (10), 116–120. doi:10.1145/570907.570945

Shneiderman, B., Fischer, G., Czerwinski, M., Resnick, M., Myers, B., Candy, L., et al. (2006). Creativity Support Tools: Report from a U.S. National Science Foundation Sponsored Workshop. *Int. J. Human-Computer Interaction* 20 (2), 61–77. doi:10.1207/s15327590ijhc2002_1

Shorofat, A. (2007). *The Effect of Using the Brainstorming Strategy on Developing Ninth Grade Students' Creative Writings Skills in Arabic Language*. Amman, Jordan: The University of Jordan.

Silvia, P. J. (2011). Subjective Scoring of Divergent Thinking: Examining the Reliability of Unusual Uses, Instances, and Consequences Tasks. *Thinking Skills and Creativity* 6 (Issue 1), 24–30. doi:10.1016/j.tsc.2010.06.001

Smith, G. J. W., and Carlsson, I. (1983). Creativity in Early and Middle School Years. *Int. J. Behav. Dev.* 6 (2), 167–195. doi:10.1177/016502548300600204

Subiaul, F., and Stanton, M. A. (2020). Intuitive Invention by Summative Imitation in Children and Adults. *Cognition* 202, 104320. doi:10.1016/j.cognition.2020.104320

Torrance, E. P. (1966). *Torrance tests of creative thinking—norms technical manual research edition—verbal tests, forms A and B—figural tests, forms A and B*. Princeton: Personnel Pres. Inc.

Torrance, E. P. (1967). *Understanding the fourth grade slump in creative thinking. Final report*. ERIC.

Torrance, E. P. (1968). A longitudinal examination of the fourth grade slump in creativity. *Gifted Child Quart.* 12 (4), 195–199.

Urban, K. K. (2005). Assessing Creativity: The Test for Creative Thinking-Drawing Production (TCT-DP). *Int. Educ. J.* 6 (2), 272–280.

Westlund, J. M. K., Park, H. W., Williams, R., and Breazeal, C. (2018). "Measuring Young Children's Long-Term Relationships with Social Robots," in IDC '18: Proceedings of the 17th ACM Conference on Interaction Design and Children (Trondheim, Norway: ACM), 207–218. doi:10.1145/3202185.3202732

Whiten, A., McGuigan, N., Marshall-Pescini, S., and Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philos. Trans. Roy. Soc. B: Biol. Sci.* 364 (1528), 2417–2428.

Williams, R., Park, H. W., Oh, L., and Breazeal, C. (2019). "Popbots: Designing an artificial intelligence curriculum for early childhood education," in Proceedings of the AAAI Conference on Artificial Intelligence, 9729–9736.

Witt, L. A., and Beorkrem, M. N. (1989). Climate for creative productivity as a predictor of research usefulness and organizational effectiveness in an R&D organization. *Creativity Res. J.* 2 (1-2), 30–40.

Yando, R., Seitz, V., and Zigler, E. (1978). *Imitation: A developmental perspective*. Lawrence Erlbaum.

Zheng, W., Wang, M. L., and Yin, J. (2013). "Correlation analysis of scaffolding creative problem solving through question prompts with process and outcomes of project-based service learning," in Proc. ASEE Annu. Conf. Expo. 1–19.

Conflict of Interest: The reviewer (KF) declared a past co-authorship with one of the authors (CB) to the handling Editor.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 12 October 2021
doi: 10.3389/frobt.2021.674121

The Sounds of Softness. Designing Sound for Human-Soft Robot Interaction

Jonas Jørgensen and Mads Bering Christiansen*

Center for Soft Robotics, SDU Biorobotics, University of Southern Denmark, Odense, Denmark

In this article, we report on research and creative practice that explores the aesthetic interplay between movement and sound for soft robotics. Our inquiry seeks to interrogate what sound designs might be aesthetically engaging and appropriate for soft robotic movement in a social human-robot interaction setting. We present the design of a soft sound-producing robot, SONŌ, made of pliable and expandable silicone and three sound designs made for this robot. The article comprises an articulation of the underlying design process and results from two empirical interaction experiments ($N = 66$, $N = 60$) conducted to evaluate the sound designs. The sound designs did not have statistically significant effects on people's perception of the social attributes of two different soft robots. Qualitative results, however, indicate that people's interpretations of the sound designs depend on robot type.

Keywords: soft robotics, human-robot interaction, sound design, social robotics, practice-based artistic research

OPEN ACCESS

Edited by:
Patricia Alves-Oliveira, University of Washington, United States

Reviewed by:
Yuhan Hu, Cornell University, United States
Yomna Abdelrahman, Munich University of the Federal Armed Forces, Germany

***Correspondence:**
Jonas Jørgensen
jonj@mmmi.sdu.dk

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *28 February 2021*
Accepted: *11 August 2021*
Published: *12 October 2021*

Citation:
Jørgensen J and Christiansen MB (2021) The Sounds of Softness. Designing Sound for Human-Soft Robot Interaction. Front. Robot. AI 8:674121. doi: 10.3389/frobt.2021.674121

INTRODUCTION

Both in real life and in science fiction movies, there exist several examples of different nonverbal and nonlinguistic sounds that robots emit as they move about or manipulate objects. Often these sounds are of a mechanical character and result from e.g. the rotations of an electrical motor, the grinding of metal parts in a joint or in a linear actuator, or the hydraulic extension of a piston. Within the cultural imaginary, robotic sounds resulting from actuation and movement thus arguably comprise their own separate category with certain established expectations and conventions associated. But what happens if the functional rigid mechanical parts responsible for these emissions of sound are replaced by pliable and soft components?

In the past two decades, *soft robotics* has become a rapidly expanding research field with an increasing number of publications each year (Bao et al., 2018). Soft robotics research seeks to replace conventional components used for building robots with pliable and elastic ones, to gain functional advantages such as energy efficiency, increased maneuverability in unstructured environments, and increased safety through passive compliance for tasks that require close human-robot interaction (Abidi and Cianchetti, 2017; Luo et al., 2017; Santina et al., 2017; Wang et al., 2017).

At present, most soft robots are pneumatically actuated with electrical pumps or compressors, but actuators without mechanical sound based on e.g. dielectric elastomers, shape memory alloys and polymers, or biological cells are gradually becoming more common (El-Atab et al., 2020; Walker et al., 2020). Hence, in the future, soft robots may become practically devoid of sound. With which sounds should a soft robot's movements and actions then be made audible to ensure safe, intuitive, and enjoyable interactions with humans?

In this article, we report on research and practice that explores the interplay between movement and sound in relation to how people experience a soft robot. More specifically, our inquiry seeks to interrogate what sound designs might be aesthetically engaging and appropriate for soft robotic movement within a social human-robot interaction setting. We present the design of a soft sound-producing robot made of pliable and expandable silicone and methods that we have used to design sound for soft robots anchored in practice-based artistic research. The article comprises an articulation of the underlying design process and two empirical experiments that examine what effect different sound designs have on people's social perception of two different types of soft robots.

This article thus addresses the following three research questions:

- RQ1: What does a soft robot sound like and what is "soft" sound?
- RQ2: What effect does "soft" sound have on people's social perception of a soft robot?
- RQ3: Are "soft" sounds a more appropriate match for a soft embodiment?

In relation to the wider theme of this special issue, the article contributes methodologically by illustrating and detailing how creative approaches and artistic methods can be integrated into human-robot interaction (HRI) research and contribute to articulating other questions and provide paths to novel insights. In addition, it presents a technical system designed to generate sounds to accompany soft robotic movement as a means of nonverbal signaling to human users. Finally, we report the results from a user study conducted to shed light on how sound affects people's assessment of a soft robot's sociality.

RELATED WORK

We position the work in the context of research on soft robotics, human interaction with soft robots, and sound design for robots.

Soft robots can be defined as systems that are capable of autonomous behavior and primarily composed of materials with elastic moduli in the range of that of soft biological materials (Rus and Tolley, 2015). Soft robots are claimed to offer inherently safer interactions with humans (Laschi et al., 2016), yet only a few publications have addressed how humans experience soft robots and how intuitive and engaging human interaction with them might be designed (Jørgensen, 2017a; Zheng, 2017; Boer and Bewley, 2018; Hu et al., 2018; Jørgensen, 2018; Shutterly et al., 2018; Zheng, 2018; Milthers et al., 2019; Jørgensen and Ploetz, 2020; Jørgensen et al., 2021; Zheng and Walker, 2019). Soft robotics technology has recently made its way into art, design, and architecture projects (Jørgensen, 2017b; Jørgensen, 2019). Yet adding sound to soft robots has not been explored within academic research and, to the best of our knowledge, only once within another creative practice project (Budak et al., 2016).

Sound has been argued to be a vital element of human communication and interaction, which should be supported in HRI. A number of HRI publications have called for more focus on sound, but robot sound design is still a nascent field of research. The addition of sound to robots has been argued to potentially improve human communication with robots and allow for more complex and meaningful interactions (Duffy, 2003; Cha et al., 2018; Jeong et al., 2017). Sound signals may also be more effective than visual cues for conveying emotional states in social robotics (Jee et al., 2009) and in HRI sound can be used to engage, inform, convey narratives, create affect, and generate attention (Schwenk and Arras, 2014). Research on robot sound design has taken many different forms including the voice-based teacher robot, *Silbot* (Jee et al., 2010), interactive sound generation with the humanoid *Robot Daryl* (Schwenk and Arras, 2014), Breazeal's sociable infant robot *Kismet* with childlike sounds (Breazeal, 2002), as well as studies investigating people's aural impressions of servo motors (Moore et al., 2017).

While many research efforts have centered on recreating human or animal sounds and human speech artificially (Duffy, 2003), recent research also exists that challenges this approach. It has been argued, for instance, that mimicking human or animal sounds could raise false expectations about a robot's abilities (Schwenk and Arras, 2014).

Prior studies on *nonlinguistic utterances* (*NLUs*) as communicative and affective means of social robotics (Read and Belpaeme, 2010; Read and Belpaeme, 2012; Read and Belpaeme, 2013; Read and Belpaeme, 2014a; Read and Belpaeme, 2014b; Rosenthal-von der Pütten and Straßmann, 2018; Wolfe et al., 2020) have used highly varied sets of discrete machine-sounding audio cues, similar to the blips and bleeps of robots in sci-fi movies. Unlike these, the sound-producing system we use here was designed to generate a coherent soundscape to accompany and augment the robot's movements and behaviors. Moreover, albeit using synthesizers for audio generation, our sound designs were purposely designed to embody both organic and machine-like qualities, and in these respects differ from research on NLUs (see *Design* detailing the design).

MATERIALS AND METHODS

Methodology

We designed and fabricated a custom pneumatically actuated soft robot, *SONŌ* (**Figure 1**), and set up two interaction experiments that investigate how different sound designs influence people's perception of a soft robot's social attributes.

The design of the SONŌ robot and its sound is anchored in practice-based artistic research drawing on both authors' practices within robotic art, electronic music, and sound design. Artistic research has, within the past two decades, been theorized as a specific mode of knowledge production. It can broadly be described as research in and through art practice that seeks to make present and communicate aesthetic experiences gained in creative practice and embodied in artistic products

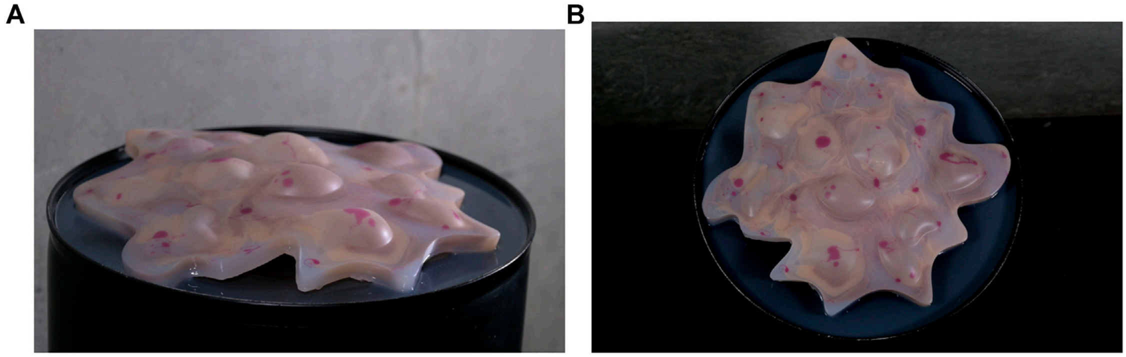


FIGURE 1 | The SONŌ robot. Side view **(A)** and top view **(B)**.

(Borgdorff, 2010). Artistic research diverges from artmaking in general, as it encompasses an ambition to contribute toward thinking and understanding and not just the development of an art practice in itself. It is linked to and engages with wider research communities, areas, or issues and hence by definition entails more than just the production of artworks. Methodologically, artistic research differs from traditional types of academic research in a number of respects. For instance, the requirement that a research study sets out with well-defined questions, topics, and problems, is at odds with the experimental character of art. Artistic research is instead undertaken on the basis of intuition, guesses, and hunches and is characterized by being open to serendipitous discoveries made along the way. Moreover, the exploration and navigating of unknown aesthetic and conceptual territories is facilitated by tacit understandings, accumulated experience, and artistic sensitivities rather than by pursuing answers to explicitly stated, rigorous, and unambiguous research questions *via* formalized methods. Hence, artistic research is discovery-led and not hypothesis-led in character (Borgdorff, 2010; Borgdorff, 2013).

We utilized practice-based artistic research methodology to address RQ1 ("What does a soft robot sound like and what is "soft" sound?"). We conducted an empirical user study, using established human-robot interaction methods and tools to evaluate the artistic outcomes in the context of HRI research and answer RQ2 ("What effect does "soft" sound have on people's social perception of a soft robot?") and RQ3 ("Are "soft" sounds a more appropriate match for a soft embodiment?"). The article thus extends prior work that has studied or evaluated robotic artworks and robot prototypes made by artists through empirical HRI experiments and prior work on leveraging the embodied meaning-making skills of artists to design robots (Demers, 2014; Vlachos et al., 2016, 2018; Levillain et al., 2017; LaViers et al., 2018; Cuan et al., 2018a; Cuan et al., 2018b; Gemeinboeck and Saunders, 2018; Gemeinboeck and Saunders, 2019; Herath et al., 2020).

Design

The practice-led research started out from the speculative question "What does a soft robot sound like?". Our intention was to experiment with how incorporating sound into a soft robot could add to its qualities and to explore how sound might support the inherent aesthetic qualities of soft robotics technology [early work has previously been reported in a Late-Breaking Report and a video (Bering Christiansen and Jørgensen, 2020a; Bering Christiansen and Jørgensen, 2020b)]. As research shows people to have a better impression and understanding of products and designs where two or more sensuous modalities are coupled (Langeveld et al., 2013), we chose to focus on how sound might augment soft robotic movement.

Design and Fabrication of the SONŌ Robot

In our design of the robot morphology we aimed for a design that would be perceived as organic yet unfamiliar. We chose a non-anthropomorphic and non-zoomorphic form and used abstract rounded shapes and a main color similar to Caucasian skin with reddish colorations to give the robot organic connotations. We opted for a simple design with only three independent pneumatic channels that each connect 4 chambers that can expand upon inflation and are located across the morphology (**Figure 2**). We deemed this design to provide sufficient possibilities for variation in realizable expressive movement.

The soft morphology was cast from Ecoflex 00-30 silicone colored with Silc-Pig pigments in a 3D printed mold (**Figure 2B**), using the following fabrication procedure. Three different containers with liquid silicone were mixed and degassed in a vacuum chamber. The first contained a light Caucasian skin tone-like pigment, the second a delicate pink pigment, and the third uncolored semitransparent silicone. The three liquid silicones were mixed directly inside the mold. A coloring with similarities to the faux marble paint effect was created by switching between the three liquid silicones when pouring them into the mold. Finally, smaller dots of deep red/purple pigmented silicone were dripped into the uncured silicone surface from a 20 cm distance with a small brush. The cured top part was removed from the mold and cast onto a strain limiting bottom piece consisting of precured silicone-coated nonwoven mesh (Vlieseline S13). Finally, three transparent supply tubes in PVC with a length of 90 cm each and 1.5 mm/3 mm ID/OD were inserted into each pneumatic channel of the soft morphology from below and the robot was coated with talc powder to prevent lint and dust from adhering to it.

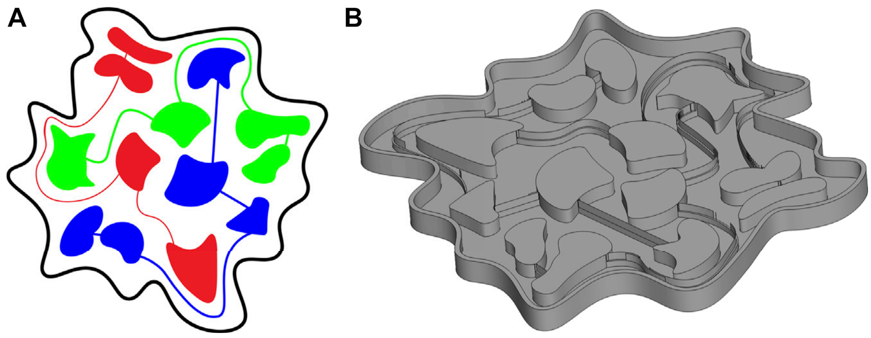


FIGURE 2 | SONŌ air chamber overview **(A)** and CAD rendering of the mold **(B)**.

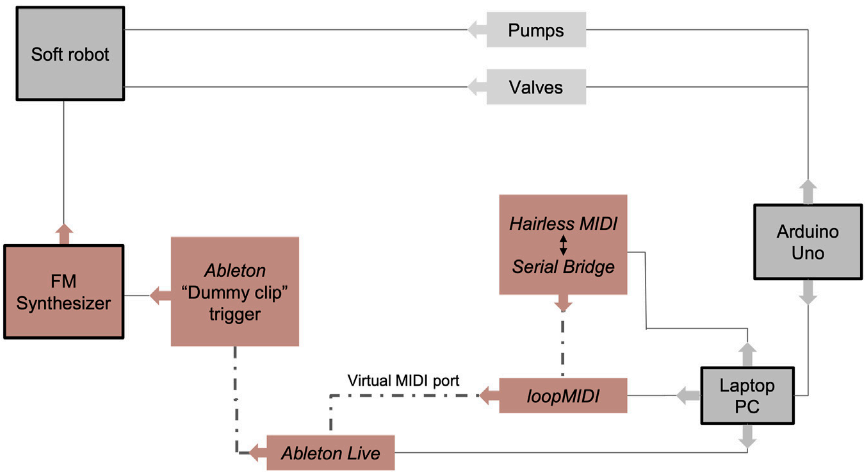


FIGURE 3 | Diagram of the technical system. Gray boxes indicate physical elements and red boxes software applications and protocols.

System Overview and Technical Setup

Figure 3 shows an overview of the system. The three physical main components are the soft robot morphology, an Arduino UNO microcontroller, and a laptop PC with a connected active speaker. The Arduino is equipped with a custom-made motor shield that drives three low noise pumps (MITSUMI R-14 A213) and three solenoid valves (Uxcell Fa0520D 6V NC) to control the soft robot's inflation and release of air.

The SONŌ robot does not currently possess any sensors for feedback control. It uses an open-loop control and switches between preprogrammed movement sequence that are executed by activating the three pumps and three valves with manually programmed time delays (Arduino code available as **Supplementary Material**). This creates bulges on the top part where the compartments are found. Expressive movement primitives of the morphology were discovered empirically through aesthetic experimentation with the robot and were combined to form programmed movement sequences. In parts of the movement sequences only one pneumatic channel is actuated, whereas in others two or all three channels inflate or deflate simultaneously. The movements performed by this and the other robot used for the interaction experiments are demonstrated in the accompanying video (Youtube link: https://youtu.be/vKHTJe8t-R0).

Frequency Modulation Synthesis

We used *frequency modulation* (*FM*) synthesis to design sound for the robot due to this technique's customizability and malleability and because it is argued to recreate natural sounds better than other forms of analogue synthesis (Jenkins, 2007). FM synthesis is based on pitch modulation of one or more oscillators (Jenkins, 2007). An FM synthesizer consists of *operators*, a term used to describe individual oscillators with separate amplitude envelopes. The amplitude of one or more *modulator operators* affects the frequency of the *carrier operator* through an algorithm, i.e. the configuration of how multiple operators interact. Depending on the algorithm, an operator can modulate other operators, be modulated by other operators or both, which has a substantial effect on the synthesizer's sound and timbral qualities with no use of filters. With FM synthesis it is possible to generate sound designs with rich complex harmonics that are impossible to create with other synthesis techniques (Chowning and Bristow, 1987).

Audio Generation

Audio to accompany the robot's movements is generated in real time by the software synthesizer *Operator* running within the *digital audio workstation (DAW)*, *Ableton Live*, on a laptop computer. The microcontroller sends MIDI signals to the DAW by utilizing a serial connection-to-MIDI-bridge, *Hairless MIDI*, and a virtual loopback MIDI-port, *LoopMIDI* (detailed setup guide included as **Supplementary Material**). The MIDI signals are sent via serial connection over USB when a pump or valve is switched on which triggers a note on the FM synthesizer. When an air chamber inflates, the frequency of the carrier operator and modulator operator(s) increases, and it decreases when an air chamber deflates. In the current setup, the robot switches between preprogrammed movement sequences, and accompanying MIDI messages sent from the microcontroller trigger "MIDI dummy clips." Dummy clips are silent MIDI clips within Ableton Live that contain an automation for modulating certain parameters of one or more devices—in this case the Operator FM synthesizer's oscillator and filter cutoff parameters. Different dummy clips have been created to contain actions that fit both inflation and deflation of the soft robot: if an air chamber inflates, a dummy clip containing inclining oscillator curve manipulation is triggered, and if an air chamber deflates, another clip containing declining oscillator curve manipulation will play. Multiple dummy clips to each sound design have been added to the DAW to allow for a less static sound image. The dummy clips have different lengths and are triggered selectively in the microcontroller control code so that they match the time a specific inflation or deflation takes, i.e. the sound does not stop abruptly.

Three sound designs were made as individual *patches*, preconfigured combinations of oscillators, filters, and envelope settings (Roland, n.d.), for the FM synthesizer. Technical details on the three patches and Ableton patch files are included as **Supplementary Materials**.

First Sound Design: "Movies"

A sound's identity—its spectro-temporal characteristics such as pitch, timbre, duration, and level—and the location of its source allows people and animals to extract relevant information from audio (Carlile, 2011). Auditory perception relies on information derived from these features that is recombined in the brain into useful and decodable signals (Carlile, 2011). Every sound and acoustic event can be understood as a decodable sign carrier that communicates information (Jekosch, 2005). For living creatures, a distinction can be made between *internal* and *external auditory cues* (Cha et al., 2018). Internal auditory cues are sounds entirely generated by the creature's own body such as breathing, snoring, or sighing, while external auditory cues are produced by its physical interaction with the environment. Echoing this distinction, commercial sound designers differentiate between *consequential sounds* and *intentional sounds* (Langeveld et al., 2013). Consequential sounds occur due to the mechanical functioning of a product's parts, intentional sounds are auditory instances meant to be triggered when products interact with their surroundings (Langeveld et al., 2013). Consequential sounds, e.g. actuation sound coming from electrical motors, are often regarded as noisy and are restricted by the physical design and properties of the product. Intentional sounds, on the contrary, are deliberate and designed.

As we did not want our initial sound design to directly mimic animal and human sounds, we started out by studying sounds made by imaginary soft characters portrayed in movies. We sought to familiarize ourselves with this existing pop-cultural frame of reference, to gain an understanding of what soft entities have been imagined to sound like and to attain insight into how their sound designs have been generated. We chose this approach as we reasoned that aligning our "soft" sound design with the formal traits of this existing repertoire of "soft" sounds could yield recognizability and make the listener associate the sound design with (fictional) soft beings. Initially, movies that contain characters with soft bodies and/or morphing/deformable soft tissue were identified by searching the internet and going through user lists on the online movie database *IMDb* (https://www.imdb.com/). Summaries and trailers for relevant movies were screened and based on this process we identified 10 movies wherein sound was a prominent feature of a soft imaginary character [*Alien* (Scott, 1979), *Alien vs. Predator* (Anderson, 2004), *Flubber* (Mayfield, 1997), *Night of the Creeps* (Dekker, 1986), *Slither* (Gunn, 2006), *Spiderman 3* (2007), *Terminator 2* (Cameron, 1991), *The Blob* (Yeaworth and Doughten, 1958), *The Thing* (Carpenter, 1982), and *Venom* (Fleischer, 2018)]. Both authors studied clips of each of these movie characters and wrote notes on what characterized their sound and how it changes upon interaction, differentiating between internal and external auditory cues. We discussed these notes and mapped shared defining features of the characters' sounds that could be considered vectors spanning the sound design space of the soft movie characters. We made the following general observations:

- Sound is dynamic (there are often rapid changes in the sound)
- Two strategies for generating sound are prevalent: 1. Recorded sounds from animals are layered, 2. Layered sounds from synthesizers are used
- Sounds are often manipulated by raising or lowering pitch or using filters. This creates "wet" or "slippery" sounds, which change in accordance with the character's movements
- Internal auditory cues convey the character's state of mind and mood. External auditory cues provide information concerning the character's movements and physical interaction with the environment

Utilizing the above observations as design guidelines, we created the SONŌ robot's first sound design patch named "Movies." The patch uses two square wave modulator operators, whose frequencies are modulated in opposite directions through a dynamic lowpass filter when movement occurs. The FM synthesizer is routed through a virtual tape echo delay with a short delay time and a high feedback percentage, which results in frequency fluctuations and overall frequency manipulation.

Second Sound Design: "White Noise"

For the second sound design, we wanted to design a "soft" sound using a different approach. We started by discussing what we understand a "soft" sound to be, in order to articulate our accumulated tacit understandings gained through experience in creative sound practice, and came to agree on some general characteristics (long envelope, timbre/spectrum not high-pitch, gradual/slow changes, lowpass filter smoothing). From this starting point, we further researched how "soft" sound is described in the literature.

In relation to sound, there are different ways in which "soft" can be understood. Dictionaries describe "soft" sound as "quiet in pitch or volume" (Merriam Webster, 2021), "gentle" and "not forceful" (Cambridge Dictionary, 2021), something "not harsh," or "low and pleasing" (Collins Dictionary, 2021). Within psychoacoustics, loudness is understood as an attribute of auditory sensation ranging on a scale from "soft," which describes low amplitude sounds, to "loud," which describes high amplitude sounds (Lentz, 2020; Fastl and Zwicker, 2007). In relation to pitch and timbre, a "soft" sound is usually low-frequency (Cook, 1999) and has less brightness than a loud sound (Cook, 2011). However, simple tones with no timbral harshness and a lack of power in the low-frequency domain, such as sine tones, have equally been described as "soft" (Howard and Angus, 2009; Seidenburg et al., 2019). The word "piano," which translates as "soft," is also used within music theory to describe a decrease in a musical score's intensity achieved by playing an instrument more gently, whereby not only the sound's amplitude is changed, but also its timbral qualities (Cook, 1999).

As the above usages of the word "soft" illustrate, different meanings persist that each point to different physical characteristics of soft sound. In our design of the second "soft" sound we chose to disregard soft as the opposite of loud, and instead focus on soft sound as the opposite of hard or harsh sound.

The patch for the second "soft" sound, named "White Noise," is based on a white noise signal. It produces a fizzing high-pitch sound with a dynamic filter cutoff. This gives the patch similarities to natural sounds such as wind or ocean waves. Based on the movements of the soft robot, the filter cutoff frequency, filter envelope percentage, and filter end position percentage are modulated in the sound design.

Third Sound Design: "Glass Attack"

As the third sound, we wanted to make a "hard" sound to contrast and compare the two "soft" sounds against. We came up with the idea to construct a sound design which sounds similar to a sound that can be produced by a hard object.

The third patch is a midtone sine wave with a relatively short amplitude attack time that includes many high-frequency harmonics, which contribute to a bell-like or glass-like sound, hence we gave it the name "Glass Attack." It sounds somewhat similar to the sound emitted by a drinking glass when the glass is brought to resonate by gently rubbing a wet finger along the rim of the glass. When the robot moves, the synthesizer uses gliding pitch manipulation to indicate inflation and deflation. This produces a sound with similarities to the resonance or impact sounds of objects made from glass or metal, with a gliding pitch manipulation added to prevent the sound from becoming static.

User Studies

We conducted two interaction experiments to test the impact of the sound designs on people's social perception of a soft robot. In the first (Experiment 1) the three sound designs were tested on the SONŌ robot, and in the second (Experiment 2) they were tested on another soft robot (**Figure 4**). This second robot is a pneumatically actuated soft silicone tentacle hanging from an aluminum frame, which was used in another study (Jørgensen et al., 2021). A four fingered soft robotic pneunets gripper cast in Ecoflex 00-30 was added to the tip of this tentacle (Finio, 2013). In the following, we will refer to this robot as the *Tentacle* robot.

We chose to test the sound designs on two different soft robot types to gain insight into whether the sound designs had similar effects, when used on soft robots in general, or if there were differences related to the type of robot using them. Both experiments had three conditions corresponding to the three sound designs, and in each experiment each robot performed the same preprogrammed movement sequence in every experiment condition (only the sound differed). We used a between-subjects design to gauge people's first impressions of the robots and avoid bias due to carry-over effects. The experiments took place over 4 days at the University of Southern Denmark (Odense) in a classroom in the main university building.

Participants

Participants were a convenience sample of people present, of whom all but one participant turned out to be university students. Demographic information for each condition is given in **Table 2** and **Table 3** under *4 Results*. Participants did not receive any compensation for their participation. Experiment 1 had a total of 66 participants and Experiment 2 had 60 participants.

Data Collection

We used the Robotic Social Attributes Scale (RoSAS) to measure people's impressions of the robot's sociality and added additional questions to obtain information about their perception of the robot's sound. The RoSAS scale is a validated tool that can be used to measure people's impressions of a robot's sociality (Carpinella et al., 2017). The scale measures three main constructs, with 6 subitems each: *Competence* (*Reliable, Competent, Knowledgeable, Interactive, Responsive, Capable*), *Warmth* (*Organic, Sociable, Emotional, Compassionate, Happy, Feeling*), and *Discomfort* (*Awkward, Scary, Strange, Awful, Dangerous, Aggressive*). Participants were asked: "Using the scale provided, how closely are the words below associated with the robot you have just experienced?". Ratings were given on a 7-point scale (1—not at all, 7—very much so). We aimed to have at least 20 participants

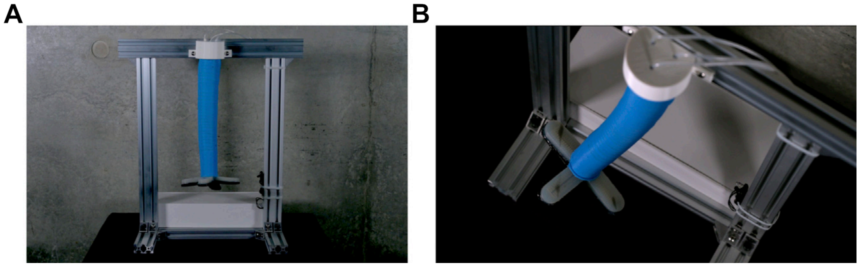


FIGURE 4 | The tentacle robot. Front view **(A)** and top view **(B)**.

per condition as other studies using the RoSAS scale with a similar number of conditions have found this to give sufficient statistical power (Pan et al., 2018).

Before filling out the RoSAS scale, participants were asked to "Write the first three words that come to mind to describe the robot that you have just experienced," following the method proposed by Damholdt et al. (2019). Another scale question was added after the RoSAS scale: "Using the scale provided, please indicate to which extent you agree with the following statement about the robot: 'The robot has a sound that is appropriate for it'" (1-Strongly agree, 7-Strongly disagree). This question was followed by an open question asking people to elaborate on their choice of answer.

As the experiments took place on the campus of a Danish university, we translated the questionnaire into a Danish version. We pretested the Danish questionnaire with five participants who experienced a video equivalent of condition 1 of Experiment 1. We changed two translated words ("kapabel" to "duelig," "responsiv" to "reaktionsdygtig") that participants expressed difficulty in understanding.

Procedure

The procedures for Experiment 1 and Experiment 2 were identical, only difference being the robot used and the pre-experiment briefing given to participants.

We asked people if they would like to participate in a research study on human-robot interaction, and upon acceptance they were accompanied to the classroom where the experiment would take place. Participants received information about the project and the experiment verbally and were given an information sheet and provided opportunity to ask questions. We did not specify to participants that the study's focus was on sound, to avoid a bias in the RoSAS ratings of the robots, as the RoSAS scale is designed to assess the overall sociality of a robot and not a single aspect of it such as its sound. Withholding this information was approved by the university's ethics committee, on the condition that it be provided in the debriefing. Written informed consent was obtained from participants, who were all above the Danish legal age of 18, both for participation in the experiment and for the collection of personal data.

In the experiments, one of the robots was placed on a table covered in dark gray cloth (**Figure 5**). In Experiment 1 that used the SONŌ robot, the electro-pneumatic actuation and control system was hidden underneath the table inside a small enclosure made from mattress foam (to dampen mechanical sound), and only the soft morphology was clearly visible on the table. The Tentacle robot instead had the actuation and control system hidden inside an integrated white acrylic enclosure (**Figure 4**).

FIGURE 5 | Photo of the experiment settings.

TABLE 1 | Average and maximum sound levels (in decibel) for the sounds occurring in the two experiments. The sound levels of the sound designs have been measured without the robot moving. The mechanical sound of the robots themselves has been measured without any sound playing on the loudspeaker. The measurements were done with the robots and the loudspeaker positioned on the table and cloth as described above. A sound meter was positioned at a fixed distance of 1 m to the robot corresponding to the approximate distance and height at which the sound would be heard by a participant.

	Avg	**Max**
1-SONŌ-White (only sound design)	66.4	77.6
2-SONŌ-Glass (only sound design)	57.0	68.6
3-SONŌ-Movies (only sound design)	60.1	72.9
4-Tent.-White (only sound design)	63.0	71.2
5-Tent.-Glass (only sound design)	59.3	73.5
6-Tent.-Movies (only sound design)	58.5	68.2
SONO robot (no sound design)	43.7	56.9
Tentacle robot (no sound design)	38.7	45.1

For sound, we used a portable active speaker, which was placed under the table for the SONŌ robot and on the table behind the Tentacle robot (**Figure 5**). We adjusted the volume of the speaker to compensate for the dampening of the sound when placed under the table, so that the loudness of each sound design was experienced as approximately the same in the two experiments. We also made adjustments to ensure that the three sound designs were experienced as being of approximately the same loudness, with sound levels as given in **Table 1**. As can be seen from **Table 1**, the mechanical sounds produced by the robots themselves were lower than the sound designs playing on the loudspeaker and we estimate that they were barely noticeable during the experiments.

We instructed participants that we would show them a robot, and that they should observe it and let us know when they felt ready to answer some questions about it. Participants experienced the robot individually or in groups of up to five persons for 2–6 min. Groups were formed as we recruited passersby on a hallway, where people were often walking in groups. We allowed a group of people to enter the premises together when several volunteered to participate at the same time. Owing to the Covid-19 pandemic we could not allow physical interaction with the two robots, hence the robots performed preprogrammed movements accompanied by sound and did not respond to participants. We encourage the reader to consult the accompanying video to get an impression of the robots and the three sound designs (Youtube link: https://youtu.be/vKHTJe8t-R0). For the SONŌ robot, participants were additionally asked to imagine that the robot was communicating and expressing itself through its movements and sound. We did this to frame this robot as a social robot. For the Tentacle robot, we instead briefed people to imagine that they were to solve a practical task, such as packing or moving small goods or products, together with this robot. We also explained that the attached soft gripper could grasp and pick up objects. We did this to frame the Tentacle robot as a soft collaborative robot (cobot).

Following exposure, participants filled out the questionnaire and provided selected demographic data (age, gender, level of familiarity with robots, prior human-robot interaction experience, field of study if a student at the university). Participants could choose freely between the Danish and the English version of the questionnaire, with 110 choosing the Danish version and 16 the English version. Finally, participants received a debriefing and were provided the opportunity to ask questions.

Hypotheses

We hypothesized that a sound design with "soft" qualities would:

- Result in higher *warmth* and *competence* ratings and lower *discomfort* ratings, than one without these qualities.
- Be deemed more appropriate for a soft robot.
- Elicit word associations with a higher rate of positive sentiments.

RESULTS

Robotic Social Attributes Scale Ratings and Appropriateness

The internal consistency of the RoSAS data was confirmed by two internal reliability tests performed on the complete data set. We calculated Cronbach's alpha, a commonly used measure of the internal consistency reliability among a group of items that form a scale. We additionally calculated the mean inter-item correlation, a more appropriate measure of internal consistency for scales with less than ten items (Briggs and Cheek, 1986). Cronbach's alpha values of 0.75 for *competence*, 0.73 for *warmth*, and 0.75 for *discomfort* were obtained, which are above the standard 0.70 threshold, indicating an acceptable internal consistency. The mean inter-item correlations were 0.33 for *competence*, 0.33 for *warmth*, and 0.34 for *discomfort* and fall within the optimal range of 0.2–0.4.

We used one-way between-groups analysis of variance (ANOVA), $\chi 2$ test for independence, and Welch test, as appropriate, to assess whether age, gender, and mean values of each the three RoSAS scale main constructs differed for the three conditions in each experiment. The same methods were used to determine if there were differences in how appropriate the sound designs were rated to be for the two robots. The results for the two experiments are given in **Table 2** and **Table 3**.

We found that with respect to familiarity with robots, participants in condition 4 differed significantly from those in conditions 5 and 6 (p = 0.050 and p = 0.005 respectively). As the assumption of homogeneity was violated when comparing mean age between conditions for both experiments, we used the Welch test for this instead of ANOVA.

We found no statistically significant differences in *competence*, *warmth*, and *discomfort* ratings between the different sound design conditions in either Experiment 1 or Experiment 2.

In secondary exploratory analyses, we compared ratings for each of the RoSAS subitems between the three sound design conditions within each experiment. For experiment 1 we found a statistically significant difference (p = 0.023) for *responsive* between condition 2 (M = 2.86) and condition 3 (M = 4.28). A statistically significant difference (p = 0.000) for *aggressive* between condition 3 (M = 3.60) and both condition 1 (M = 2.05) and condition 2 (M = 1.81) was also found. For experiment 2 we found borderline statistically significant differences for *reliable* (p = 0.052) between condition 5 (M = 4.21) and condition 6 (M = 3.05) and for *awkward* (p = 0.067) between condition 4 (M = 3.30) and condition 6 (M = 4.80). We also compared how appropriate each of the three sound designs were rated to be with the SONŌ robot and the tentacle robot respectively, using T tests and data from both the experiments. We found no significant differences (p > 0.05) despite differing mean values (**Table 2** and **Table 3**).

TABLE 2 | Results and demographic data from Experiment 1. Groups under "Faculty" indicate under which faculty the participant studies if a student at the university (HUM = humanities, NAT = natural sciences, SOC = business and social sciences, HEA = health sciences, TEC = technical sciences).

	1-SONŌ-White (*N* = 20)	2-SONŌ-Glass (*N* = 21)	3-SONŌ-Movies (*N* = 25)	N	P-value(ANOVA/χ2/Welch)
Competence	M: 3.48 SD:1.01	M: 3.00 SD:1.05	M: 3.47 SD:1.11	66	0.25
Reliable	M: 3.65 SD:1.50	M: 3.33 SD:1.32	M: 2.76 SD:1.13	66	0.08
Competent	M: 3.40 SD:1.47	M: 2.95 SD:1.66	M: 3.36 SD:1.35	66	0.56
Knowledgeable	M: 2.65 SD:1.50	M: 2.76 SD:1.76	M: 3.32 SD:1.55	66	0.32
Interactive	M: 3.50 SD:1.96	M: 2.95 SD:1.80	M: 3.56 SD:1.83	66	0.50
Responsive	M: 3.75 SD:1.68	M: 2.86 SD:1.68	M: 4.28 SD:1.75	66	0.02
Capable	M: 3.90 SD:1.52	M: 3.14 SD:1.28	M: 3.56 SD:1.45	66	0.24
Warmth	M: 3.19 SD:0.99	M: 3.14 SD:1.38	M: 3.49 SD:1.13	66	0.56
Organic	M: 5.35 SD:1.50	M: 4.67 SD:2.22	M: 4.36 SD:2.02	66	0.24
Sociable	M: 2.35 SD:1.39	M: 2.29 SD:1.38	M: 2.92 SD:1.55	66	0.27
Emotional	M: 2.75 SD:1.52	M: 3.48 SD:1.86	M: 3.44 SD:1.92	66	0.34
Compassionate	M: 2.55 SD:1.32	M: 2.52 SD:1.72	M: 2.76 SD:1.39	66	0.84
Happy	M: 2.85 SD:1.42	M: 2.43 SD:1.75	M: 3.12 SD:1.62	66	0.35
Feeling	M: 3.30 SD:1.98	M: 3.48 SD:2.09	M: 4.32 SD:1.75	66	0.17
Discomfort	M: 3.15 SD:0.98	M: 3.19 SD:1.08	M: 3.65 SD:1.14	66	0.22
Awkward	M: 3.75 SD:1.34	M: 3.29 SD:2.00	M: 3.48 SD:1.36	66	0.64
Scary	M: 3.30 SD:1.92	M: 3.43 SD:1.91	M: 3.64 SD:2.00	66	0.84
strange	M: 6.00 SD:1.56	M: 5.95 SD:1.16	M: 6.00 SD:1.35	66	0.99
Awful	M: 2.10 SD:1.65	M: 2.48 SD:1.60	M: 2.84 SD:1.60	66	0.32
Dangerous	M: 1.70 SD:1.41	M: 2.19 SD:1.44	M: 2.36 SD:1.60	66	0.33
Aggressive	M: 2.05 SD:1.43	M: 1.81 SD:1.25	M: 3.60 SD:1.68	66	0.00
Appropriateness of sound	M: 4.25 SD:1.71	M: 3.95 SD:1.63	M: 4.68 SD:1.63	66	0.33
Age	M: 23.5 SD:5.23	M: 23.9 SD:2.90	M: 22.1 SD:1.49	66	0.04
Gender (female/male)	(10/10)	(6/15)	(11/14)	66	0.35
Familiarity w. robots	M: 3.35 SD:1.73	M: 3.33 SD:1.91	M: 3.04 SD:1.97	66	0.82
Faculty (HUM/NAT/SOC/HEA/TEC)	(8/2/0/8/5)	(1/4/2/2/11)	(6/3/4/8/4)	66	—

TABLE 3 | Results and demographic data from Experiment 2

	4-Tent.-White (*N* = 20)	5-Tent.-Glass (*N* = 20)	6-Tent.-Movies (*N* = 20)	N	P-value(ANOVA/χ2/Welch)
Competence	M: 3.48 SD:1.02	M: 3.69 SD:1.07	M: 3.23 SD:0.84	60	0.34
Reliable	M: 3.50 SD:1.40	M: 4.21 SD:1.65	M: 3.05 SD:1.32	59	0.05
Competent	M: 3.55 SD:1.43	M: 4.05 SD:1.50	M: 3.45 SD:1.47	60	0.39
Knowledgeable	M: 2.60 SD:1.27	M: 3.00 SD:1.26	M: 2.60 SD:1.43	60	0.55
Interactive	M: 3.85 SD:1.60	M: 3.75 SD:2.10	M: 3.10 SD:1.29	60	0.32
Responsive	M: 3.20 SD:1.24	M: 3.45 SD:1.43	M: 3.40 SD:1.67	60	0.85
Capable	M: 4.20 SD:1.44	M: 3.90 SD:1.48	M: 3.80 SD:1.58	60	0.68
Warmth	M: 2.78 SD:1.01	M: 2.28 SD:0.68	M: 2.58 SD:1.06	60	0.23
Organic	M: 3.90 SD:2.02	M: 3.35 SD:1.95	M: 3.15 SD:1.84	60	0.45
Sociable	M: 2.50 SD:1.54	M: 1.68 SD:0.95	M: 2.25 SD:1.86	59	0.23
Emotional	M: 2.05 SD:1.32	M: 1.55 SD:1.00	M: 2.05 SD:1.47	60	0.37
Compassionate	M: 2.20 SD:1.51	M: 1.50 SD:1.00	M: 1.85 SD:0.99	60	0.22
Happy	M: 3.20 SD:1.58	M: 2.40 SD:1.73	M: 2.95 SD:2.14	60	0.37
Feeling	M: 2.85 SD:1.46	M: 3.25 SD:1.83	M: 3.25 SD:1.71	60	0.69
Discomfort	M: 2.61 SD:1.18	M: 2.73 SD:1.07	M: 3.12 SD:1.11	60	0.33
Awkward	M: 3.30 SD:1.78	M: 3.85 SD:2.23	M: 4.80 SD:2.02	60	0.07
Scary	M: 2.80 SD:1.96	M: 2.20 SD:1.85	M: 2.50 SD:1.76	60	0.60
strange	M: 4.15 SD:1.95	M: 4.80 SD:1.99	M: 5.00 SD:2.10	60	0.39
Awful	M: 2.10 SD:1.45	M: 2.35 SD:1.69	M: 2.80 SD:1.64	60	0.38
Dangerous	M: 1.75 SD:1.12	M: 1.50 SD:0.83	M: 1.70 SD:1.34	60	0.76
Aggressive	M: 1.55 SD:1.00	M: 1.70 SD:0.33	M: 1.90 SD:1.21	60	0.68
Appropriateness of sound	M: 4.30 SD:1.92	M: 4.58 SD:1.81	M: 3.75 SD:2.12	59	0.41
Age	M: 22.6 SD:1.76	M: 23.0 SD:2.20	M: 25.0 SD:4.95	59	0.14
Gender (female/male)	(5/15)	(8/11)	(8/12)	59	0.47
Familiarity w. robots	M: 4.35 SD:2.16	M: 3.00 SD:1.53	M: 2.55 SD:1.47	59	0.01
Faculty (HUM/NAT/SOC/HEA/TEC)	(0/2/0/6/12)	(2/3/10/0/4)	(2/3/3/0/12)	59	—

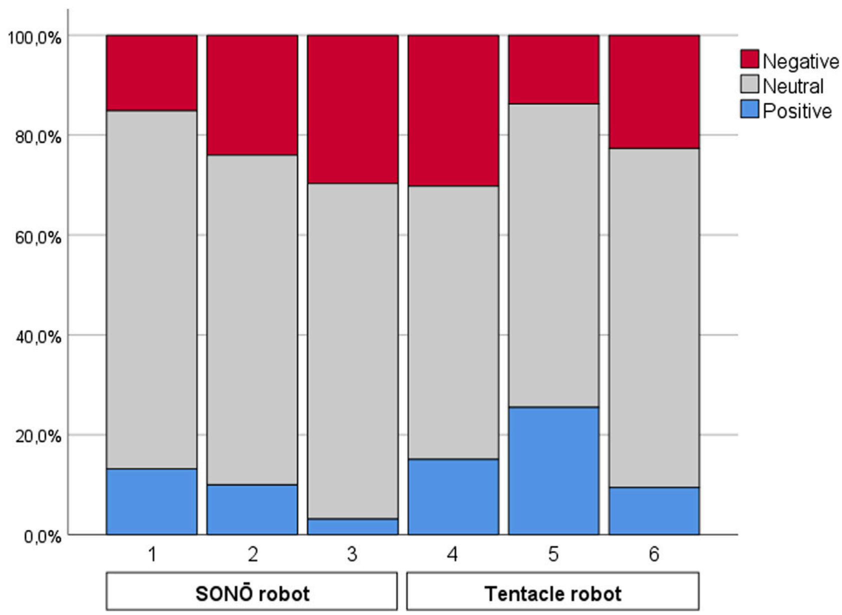


FIGURE 6 | Stacked bar graph showing proportional distributions of positive, neutral, and negative sentiment words.

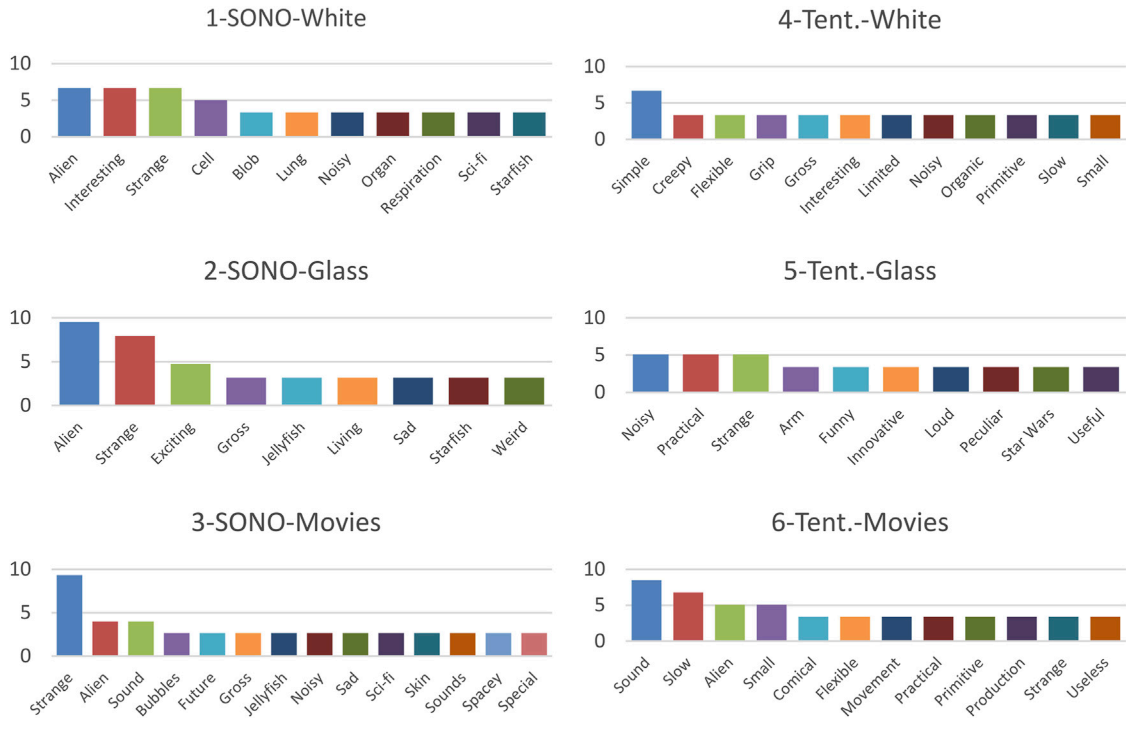


FIGURE 7 | Most frequent words used by participants. The *y*-axis gives percentage of word occurrence frequency (%).

Sentiment Analysis of Descriptive Words

Following the method described in (Damholdt et al., 2019), we used logistic regression to determine if the words used to describe the robots had different distributions of sentiment in the three sound conditions for each robot. We used this method of analysis to determine if the sound design condition could predict the likelihood that respondents would report a word with a positive sentiment.

A total of 378 responses were obtained, corresponding to 3 word entries from each of the 126 participants. The three response items from one participant who had used the three words to form one coherent entry ("star," "wars," "sounds") were reduced to two items ("star wars" and "sounds") and a blank entry, yielding a reduction to 377 items. As the next step, all Danish items were translated into English and items containing more than one word were shortened to one word, following the procedure in (Damholdt et al., 2019). Two coders then coded all words as being of either negative, neutral, or positive sentiment. Cohen's κ was run and yielded a substantial interrater reliability of $\kappa = 0.736$, the percentage of agreement was 85.9% with 324 out of 377 words categorized identically by the two coders. The identically categorized words were included for further analysis (proportional distributions for each condition are visualized in **Figure 6** and most frequent words, mentioned by two or more participants, are visualized in **Figure 7**).

Neutral words were subsequently excluded, which yielded a total of 114 either negative or positive words that were included in the logistic regression. Direct logistic regression was performed to assess the impact of a number of factors on the likelihood that a word listed by a respondent would have a positive sentiment. The model contained 5 independent variables (condition, gender, age, familiarity with robots, faculty).

For Experiment 1, the full model was statistically significant $\chi 2$ (15, $N = 53$) = 31.6, $p = 0.007$, indicating that the model was able to distinguish between words with a positive and negative sentiment. The model explained between 44.9% (Cox and Snell R square) and 65.6% (Nagelkerle R squared) of the variance in positive and negative words, and correctly classified 84.9% of cases. However, none of the independent variables made a unique statistically significant contribution to the model, with only gender ($p = 0.054$) and age ($p = 0.066$) being trend level significant.

The full model was also statistically significant $\chi 2$ (15, $N = 61$) = 31.34, $p = 0.005$, for Experiment 2, indicating that the model was able to distinguish between words with a positive and negative sentiment. The model explained between 40.2% (Cox and Snell R square) and 54.0% (Nagelkerle R squared) of the variance in positive and negative words, and correctly classified 82.0% of cases. However, none of the independent variables made a unique statistically significant contribution to the model.

Thematic Analysis

In the questionnaire, participants were asked to indicate to which extent they agreed with the following statement: "The robot has a sound that is appropriate for it." To gain further insights into people's perceptions of the different sound designs, we also asked people to elaborate on their chosen answer. We analyzed the replies by *thematic analysis*, using an inductive coding to allow for unexpected themes to emerge (Braun and Clarke, 2006).

Two coders read through all replies obtained in both experiments and respectively identified 6 and 8 codes for recurrent utterances. The coders shared their codes with each other (of which 5 were overlapping) and merged them into a coding scheme with 9 codes capable of adequately capturing and differentiating salient participant utterances. We chose to include all codes because of the thematic analysis having an exploratory aim. A codebook (included as **Supplementary Material**) was created and both coders coded the data using this coding scheme. Data items (individual participant responses) were assigned from 1 to 3 matching codes each and then exported into separate lists for each code. Finally, the first coder inductively constructed 6 recurrent themes from the lists of items, which we describe below and illustrate with exemplary quotes.

Theme #1: Loud Sound

15 participants mentioned the sound's loudness or described it as shrilling or noisy. Some ($N = 8$) explicitly expressed positive or negative opinions about the sounds. The positive comments ($N = 2$) focused on how the sound was loud but suited the robot and did not cause irritation:

"*The sound is a bit loud, but not noisy or unpleasant*" (Participant 94, 5-Tent.-Glass)

The negative comments ($N = 6$) expressed annoyance, stress, and discomfort upon experiencing the sound:

"The sound was disturbing and a bit too loud, which I felt did not suit the robot" (Participant 39, 2-SONŌ-Glass).

The majority ($N = 9$) of the 15 items in the theme come from "Glass Attack" sound design conditions [2-SONŌ-Glass ($N = 4$); 5-Tent.-Glass ($N = 5$)]. 8 of these describe this sound design as being too loud or annoying.

Theme #2: Othering Robot/Sound

Nearly a quarter of all participants ($N = 30$) described the robot and/or its sound through what we refer to as "othering." By this term we describe utterances that position the robot or its sound as something that differs from or falls outside of what is deemed to be normal and relatable. This encompasses descriptions of it as either 1) strange, weird, or mystical or 2) alien-like, science fiction (sci-fi)-like, otherworldly, or futuristic. While there were 20 instances of this for the SONŌ robot [1-SONŌ-White ($N = 5$); 2-SONŌ-Glass ($N = 8$); 3-SONŌ-Movies ($N = 7$)], there were only 10 for the Tentacle robot [4-Tent.-White ($N = 2$); 5-Tent.-Glass ($N = 3$); 6-Tent.-Movies ($N = 5$)].

12 participants described the robot or its sound as strange, peculiar, mystical, or weird and most ($N = 7$) commented negatively on this:

"The sound is strange and so is the robot. But a more pleasant sound could be more suiting" (Participant 23, 2-SONŌ-Glass).

"I found it strange, maybe unnecessary, to add an artificial sound to the actions" (Participant 90, 5-Tent.-Glass).

Only 2 participants described the "White Noise" sound design as strange [1-SONŌ-White ($N = 2$)]. The "Glass Attack" and "Movies" sound designs, on the other hand, were described as such 6 and 4 times respectively [2-SONŌ-Glass ($N = 5$); 5-Tent.-Glass ($N = 1$); 3-SONŌ-Movies ($N = 3$); 6-Tent.-Movies ($N = 1$)]. Descriptions that refer to the sound as strange principally came from Experiment 1 that used the SONŌ robot ($N = 10$), and of these, 4 additionally described the robot's appearance as strange.

18 participants instead found the robot or its sounds to be otherworldly, futuristic, or conjure up aliens or sci-fi. Some found the sound "too futuristic" (participant 99, 5-Tent.-Glass) or "overly scifi (sic) sounding" (participant 112, 6-Tent.-Movies) or stated that:

"The sound is okay, perhaps a bit UFO like" (Participant 124, 6-Tent.-Movies)

Others referenced specific examples from popular culture including the *Alien* movies. A third group pointed out that it was only some qualities of the sound that made it appear "alien" whereas other qualities contributed to the experience of the robot as being a living, autonomous, or even sentient creature.

Comments relating to sci-fi, aliens, and the future predominantly concern the "Movies" sound design [3-SONŌ-Movies ($N = 4$); 6-Tent.-Movies ($N = 4$)]. Hence, it appears that some participants were able to trace back the connection to this sound design's original sources of inspiration.

Theme #3: Associations to Other Sounds

22 participants associated the sound designs with sounds emitted from familiar man-made objects, living creatures, or natural phenomena.

Such responses for the "White Noise" sound design largely concern sounds of wind, breath, or air [1-SONŌ-White ($N = 6$); 4-Tent.-White ($N = 4$)]. One participant associated 2-SONŌ-Glass with the sound of breathing, but no other associations to wind, breath, or air were present for the "Glass Attack" and "Movies" sound designs.

Interestingly, in conditions with the "White Noise" sound design, associations differ markedly for the SONŌ and the Tentacle robot. For SONŌ the sound reminded participants of the robot breathing in sync with the robot's movements ($N = 3$). Or participants on the contrary stated that the sound of wind was not appropriate for the robot and that it instead should have had more "breathing sounds" ($N = 3$):

"*It has a breathing sound*" (Participant 4, 1-SONŌ-White)

"*My first thought was not the sound of ?wind? when I saw it*" (Participant 6, 1-SONŌ-White).

For the Tentacle robot, 1 participant argued that the "blowing sound" made the robot "more repulsive" (Participant 79, 4-Tent.-White), while another participant believed the sound to be the actual sound of the pressurized air actuating the robot and not a designed sound. Two other participants connected the sound to hydraulics, vacuum, and air-controlled machinery:

"*Because I was thinking of vacuum and the sound seems hydraulic, I find it well-suiting*" (Participant 75, 4-Tent.-White)

Some participants instead described the sound as "robot-like" ($N = 7$), mainly in the 3-SONŌ-Movies condition ($N = 4$). However, 2 participants in 5-Tent.-Glass also argued that this sound is how one would imagine a robot to sound like, and 1 participant in 6-Tent.-Movies believed that the robot "(. . .) *does not say words like humans and therefore it sounds like a robot*" (Participant 116), 6-Tent.-Movies).

While most participants did not elaborate on why or how the sound was robot-like or what a "typical" robot sounds like, 1 participant did mention specific movie examples:

"It sounds like what one always imagines a robot to sound like. Very mechanical and a sound one has heard in Star wars/ terminator [sic]" (Participant 82, 5-Tent.-Glass).

Remaining answers within the theme ($N = 5$) associated the sound with various living creatures or objects. One participant experiencing the SONŌ robot, for instance, argued that the "(. . .) *sound was stressful, sounded like a whale*" (Participant 37, 2-SONŌ-Glass). Four participants experiencing the Tentacle robot instead associated its sound with technical equipment including an airplane (Participant 87, 5-Tent.-Glass), a car (Participant 126, 5-Tent.-Glass), a robotic arm in a factory (Participant 105, 6-Tent.-Movies), and medical equipment (Participant 77, 4-Tent.-White).

Theme #4: "Organic" Appearance vs. "Mechanical" Sound

14 participants answered by evaluating the connection between the robot's visual appearance and its sound, with a majority invoking a dichotomy between organic and mechanical qualities.

For the SONŌ robot, comments ($N = 10$) predominantly described the sound as more "mechanical" or "electronic" than the "natural" or "organic" appearance of the robot ($N = 7$), and were distributed nearly evenly among the three sound designs—1-SONŌ-White ($N = 3$), 2-SONŌ-Glass ($N = 2$), and 3-SONŌ-Movies ($N = 2$):

"*It inflated with a sound that sounded more mechanical than what I would expect from something organic*" (Participant 5, 1-SONŌ-White)

"*It sounds more electronic than the organic feeling it emanates*" (Participant 59, 3-SONŌ-Movies)

Only 1 of the 4 comments for the Tentacle robot invoked a distinction between "organic" and "mechanical" (in condition 4-Tent.-White). But 7 participants commented on this for the SONŌ robot irrespective of sound design, which indicates that the SONŌ robot's embodiment or its framing as a social robot may have contributed to participants hearing the sound as mechanical.

Theme #5: Does Sound Match Appearance?

15 participants evaluated the fit between the robot's appearance and its sound for the SONŌ robot [1-SONŌ-White ($N = 5$); 2-SONŌ-Glass ($N = 5$); 3-SONŌ-Movies ($N = 5$)].

A majority of participants ($N = 9$) stated that the sound suits the robot's looks or matches what is expected for the robot's appearance, with "Glass Attack" having the highest prevalence [1-SONŌ-White ($N = 2$); 2-SONŌ-Glass ($N = 4$); 3-SONŌ-Movies ($N = 3$)]. Others (N = 6) argued that the sound design was surprising or inappropriate for the robot's appearance [1-SONŌ-White ($N = 3$); 2-SONŌ-Glass ($N = 1$); 3-SONŌ-Movies ($N = 2$)]:

"The sound's accentuated (sic) is a bit surprising compared with the robot's appearance" (Participant 62, 3-SONŌ-Movies).

"I think the sound correlates well with its appearance, sort of innocent and a bit sad" (Participant 22, 2-SONŌ-Glass).

Theme #6: Synchronized Movement and Sound

For the Tentacle robot, instead of appearance, we found a focus on the connection between movement and sound in responses. Where 6 participants commented on this connection for the SONŌ robot [1-SONŌ-White ($N = 2$); 2-SONŌ-Glass ($N = 1$); 3-SONŌ-Movies ($N = 3$)], nearly twice as many ($N = 11$) did so for the Tentacle robot [4-Tent.-White ($N = 5$); 5-Tent.-Glass ($N = 3$); 6-Tent.-Movies ($N = 3$)]. 10 of these found the sound to accompany the movements well or to be e.g. "(. . .) *in harmony with the movements*" (participant 120, 6-Tent.-Movies). The remaining participant, by contrast, described the sound as erratic, unnecessary, and not matching the robot's movements (participant 70, 4-Tent.-White). For the SONŌ robot, all 6 participants evaluated the sound designs as fitting the robot's movement:

"*There was an adequate synchronization between the movement and the sound*" (Participant 7, 1-SONŌ-White)

"It (the sound) accompanied the movements well, and it gave a sense of reliability to the machine" (Participant 97, 5-Tent.-Glass).

A possible explanation for why appearance is in focus for the SONŌ robot and movement in focus for the Tentacle robot could be that the latter has more visible movement. The tentacle changes position and bends in three dimensions, whereas the SONŌ robot's surface only inflates somewhat upward, which might to some participants not be sufficient to be regarded as "movement" and is therefore categorized as a change in the robot's appearance instead.

DISCUSSION

In this article we have explored the potentials of augmenting soft robotics with sound for human-robot interaction through the design of the SONŌ robot and its associated sound designs, and presented a system to generate sound to accompany the movements of a pneumatically actuated soft robot. Our approach was based in creative practice and artistic research methodologies combined with empirical HRI methods for testing.

Surprisingly, the quantitative results from the user study indicate that we must reject our three hypotheses; the two "soft" sound designs did not lead to higher *warmth* and *competence* ratings and lower *discomfort* ratings than the third "hard" sound design. Neither were the "soft" sounds deemed more appropriate for the two soft robots and they did not elicit a higher rate of words with positive sentiment to describe the robots.

Comparing results from both experiments, we found that there was no difference in how appropriate the three sounds designs were rated to be when comparing between the two robots, i.e. none of the sound designs were deemed a better fit for one or the other of the two robots. Hence, from this result we cannot conclude that one of the designs is especially fit for a communicative soft social robot, such as SONŌ, or a soft cobot, such as the Tentacle robot.

In exploratory analyses, however, we found that in Experiment 1, the "Movies" sound design made the SONŌ robot appear significantly more *responsive* than "Glass Attack." "Movies" also made this robot appear significantly more *aggressive* than both other two sound designs. Similarly, in Experiment 2, we obtained trend level statistically significant differences for individual RoSAS subitems: for "Movies" there was a trend toward it making the Tentacle robot less *reliable* than "Glass Attack," and more *awkward* than "White Noise." These results indicate that the sound designs do impact people's perception of very specific qualities of the robots, such as *responsiveness*, *aggression*, *reliability*, and *awkwardness*, but perhaps not the broad high-level main RoSAS constructs. Moreover, that these effects differed for the two experiments suggests that sound interacts with context or morphology in determining how specific aspects of a soft robot's sociality are evaluated.

In logistic analyses of word sentiment, we found some unexpected effects, not related to the sound designs. In Experiment 1, gender (p = 0.054) and age (p = 0.066) had marginal effects on whether a participant used a word with a positive sentiment to describe the robot. Male participants appear less likely to use a word with positive sentiment and the direction of the latter relationship matches the results obtained by Damholdt et al. (2019), where an increase in age led to a higher probability that participants would use a word with positive sentiment. This marginal effect might become significant with a wider age range, and not all but one participant being university students, as in our cohort. These effects could be interesting to study further with respect to how they compare with other effects of age on perception of and attitudes toward robots. A possible explanation for this result could be the so-called positivity effect, which describes a shift from a negativity bias in young people to a preference for positive information later in life (Carstensen and DeLiema, 2018). From **Figure 6**, which shows the proportional distributions of positive/neutral/negative words, we can equally observe a trend toward different ratios of positive-negative words in the different conditions, which might become significant with more statistical power.

In **Figure 7**, which shows bar graphs of the most used words, it can be seen that more nouns feature as recurrent descriptive words for the SONŌ robot than for the Tentacle robot, which has mostly evaluative adjectives (however, "alien" could be counted as both a verb and noun). This matches well with that the Tentacle robot was framed as a robot made for a specific practical purpose, which it is evaluated for, whereas the SONŌ robot was presented in a more open-ended scenario as a socially communicative robot.

Comparing word use between the three conditions for each robot more closely, we can see that for the SONŌ robot, the words "alien" and "strange" are both among the top three words mentioned in all conditions. Therefore, it is likely that these descriptions are independent of the three sound designs. This also matches that descriptions of this robot itself as "strange" or "alien", are prevalent in Theme 2 of the thematic analysis. Moreover, on the RoSAS scale, "strange" is rated to have a higher association with the SONŌ robot (M:5.95-6.00) than the Tentacle robot (M:4.15-5.00), and using T-tests we could verify that these differences in mean values between the robots were significant in both the "White" and "Glass" condition (p = 0.002, p = 0.032) and close to significant for "Movies" (p = 0.075). Based on this we conclude that it is likely that the SONŌ robot's embodiment or contextual framing creates the impression of the robot as being "strange" and not differences between the sound designs.

When looking at words that are used for individual sound designs for both robots, similarities are also apparent. For both conditions that used the "Movies" sound design, the word "sound" is among the two most frequently mentioned words (and the word "sounds" is additionally present for one condition). This suggests that the "Movies" sound design draws more attention to itself, than the other sound designs, which are less obtrusive and perhaps easier to integrate into the overall impression of the robot. However, either the word "noisy" or "sound" was mentioned by two or more participants in all conditions except two. That the sound of "Movies" is perceived as more dominating or assertive, aligns well with that this sound design contributed to the robot appearing more aggressive in Experiment 1.

In the thematic analysis, we uncovered 6 recurrent themes, and found differences within these in how the three sound designs were assessed qualitatively. For instance, it was predominantly the "Glass Attack" sound design that was mentioned as being loud, which the majority of participants experienced negatively (and "Noisy" and "Loud" were also among the most frequently used words to describe the Tentacle robot when it is used "Glass Attack"). Perhaps more interestingly, the thematic analysis equally showed that the three sound designs were described differently when they were used by each of the two robots. A main takeaway from the thematic analysis is thus that robot type, i.e. the robot's embodiment combined with its specified use context, appears to affect how a sound design is understood and how the specific sounds made by a soft robot are interpreted. For instance, the associations to other well-known sounds were different for the "White Noise" sound design for the two robots: While it was associated to air for both robots, for the SONŌ robot it was associated with breathing and live organisms, whereas for the Tentacle robot it was instead pneumatics and technical equipment that was mentioned. This observed difference aligns with prior work showing that embodiment affects emotional response to nonlinguistic utterances (Wolfe et al., 2020). Yet, our study design does not allow us to determine if this difference is an effect of embodiment or of context. Further work is needed to separate and distinguish between the effects of each of the two.

Returning to the research questions posed at the outset of our inquiry, "What effect does 'soft' sound have on people's social perception of a soft robot?" and "Are 'soft' sounds a more appropriate match for a soft embodiment?", the conclusion to draw is that these questions need to be asked with more nuance. Differences in sound design we authors, as creative practitioners active in the fields of electronic music and robotic art, picked up on and deemed to have a marked effect on our own perception of the soft robots, might not have enough impact on people in general, so as to make a difference with respect to how they rate impressions of high-level constructs such as *warmth*, *competence*, and *discomfort*. As we have explored through our practice-based artistic research, different kinds of "soft" sound exist and as the empirical tests showed, there were qualitative differences in how the robots were perceived when utilizing the three different sound designs. In further work, it would be relevant to study, that if different sound designs do not have marked effects on a soft robot's general sociality, then could sound perhaps affect other more basic perceived qualities of the robot? Studies have shown, for instance, that humans and animals are able to infer what material an object is made from by using visual information and impact sounds, i.e. the sound an object makes when being struck by e.g. a hammer, and that there are strong audio-visual interactions in material-category perception (Fujisaki et al., 2014). In one study, the appearance of glass combined with the impact sound of a bell pepper was thus perceived as transparent plastic (Fujisaki et al., 2014). This phenomenon is worthy of further study in relation to soft robotics, with a view to determining if sound could alter people's perceptions of a soft robot's affordances or stiffness, for instance. In a previous study (Jørgensen et al., 2021) we found that in interactions with humans, soft robots are sometimes spontaneously subjected to a more forceful handling than traditional robots, sometimes even to the point of them breaking. A possible way to prevent this, could be to add a sound to the robot that makes it appear softer or more fragile, and this way nudge the user to handle it more carefully.

As further work, we plan to develop the SONŌ robot and the system into a finished artwork and to conduct further user tests during its exhibition. This will allow us to gauge if the change of setting from a university classroom to an art exhibition contributes to different sound designs having more impact on people's perceptions of the robot, e.g. due to a heightened aesthetic awareness induced by the latter context.

LIMITATIONS

Despite offering design advantages in terms of variation, flexibility, and adaptability, FM synthesis might not be the most appropriate technique to generate "soft" sounds. Perhaps the sounds that can be created with FM synthesis are not "soft" enough, and the differences between the three sound designs are not pronounced enough to have significant effects. A limitation to the study is, that we, following common practice within artistic research, did not test whether the "soft" sounds were perceived as "soft" by lay users, or how lay users define "soft" sound, which could be done as further work. This limits the generalizability of the user study's results to the two specific definitions of "soft" sound embodied in the "Movies" and "White Noise" sound designs. Under Theme 4 in the thematic analysis, for instance, we found that for the SONŌ robot participants remark on all three sound designs that they are more "mechanical" or "electronic" than what would be expected from this robot's "organic" appearance. This could indicate that more "organic" sounds, such as e.g. recorded sounds, could be a better fit for this embodiment.

Another limitation of the user study is that the sounds which are generated by our system are synchronized with the movements being performed. Hence the sounds produced by a specific sound design change somewhat, due to varying durations between the two robots, but they do share the same characteristic overall quality. The movements of the tentacle, for instance, are based on sequences with longer inflation times; hence, the sounds emitted are also made longer with this robot.

In the thematic analysis it is evident that the "Glass Attack" sound design is perceived as loud by several participants. This could be due to that some frequencies are perceived as louder than others (Cook, 1999), and that this sound design made more use of these. To account for this, we could have asked people to rate the loudness of the sounds in a pretrial and adjusted to the perceived loudness in each condition of the experiments based on the pretrials, rather than doing this based on our own perception of the sound.

With respect to the RoSAS scale, there are several scale items that rely on interactivity, and due to the Covid-19 pandemic we were only able to display the robots to participants and the robots would not respond to them. This makes the assessments of e.g. *competence* less relevant and reliable.

A limitation could also result from the choice made to not prime participants to focus on sound. It is possible that the unfamiliar appearance of the soft robots contributed a novelty effect that trumps the effect of the sound design in the evaluations. I.e. the quaint looks of the robots might have stolen the focus from the sound and contributed to lessening the effect of differences in sound, which might have been more pronounced with a more common robot.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are available upon request. Requests to access the datasets should be directed to jonj@mmmi.sdu.dk

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee (REC) at the University of Southern Denmark. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

JJ is the main author of this article who provided the storyline and structure of the article and most of the text. JJ designed the user study and performed the statistical analyses and edited the final article. MC designed the SONŌ robot and the sound designs under the supervision of JJ. MC contributed to related work, wrote most of design, and drafted the thematic analysis section under supervision of JJ. MC shot and edited the accompanying video. Both authors contributed to setting up and conducting the user study.

ACKNOWLEDGMENTS

We acknowledge help received from Cao Danh Do with 3D prints and fabrication and from Carsten Albertsen with designing and building the electronics. We would like to thank all participants in the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2021.674121/full#supplementary-material

Supplementary Data Sheet 1 | Ableton patch files

Supplementary Data Sheet 2 | Setup Guide: Arduino/Ableton Live MIDI connection

Supplementary Data Sheet 3 | Arduino code

Supplementary Data Sheet 4 | Technical Details on FM Synthesizer Patches

Supplementary Data Sheet 5 | Code book – Thematic analysis

REFERENCES

Abidi, H., and Cianchetti, M. (2017). On Intrinsic Safety of Soft Robots. *Front. Robot. AI* 4. doi:10.3389/frobt.2017.00005

Bao, G., Fang, H., Chen, L., Wan, Y., Xu, F., Yang, Q., et al. (2018). Soft Robotics: Academic Insights and Perspectives through Bibliometric Analysis. *Soft Robotics* 5, 229–241. doi:10.1089/soro.2017.0135

Bering Christiansen, M., and Jørgensen, J. (2020a). "Augmenting Soft Robotics with Sound," in Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction HRI '20, Cambridge, United Kingdom, March 2020 (New York, NY: Association for Computing Machinery), 133–135. doi:10.1145/3371382.3378328

Bering Christiansen, M., and Jørgensen, J. (2020b). "Sonō," in Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction HRI '20, Cambridge, United Kingdom, March 2020 (New York, NY: Association for Computing Machinery). doi:10.1145/3371382.3378399639

Boer, L., and Bewley, H. (2018). "Reconfiguring the Appearance and Expression of Social Robots by Acknowledging Their Otherness," in Proceedings of the 2018 Designing Interactive Systems Conference DIS '18, Hong Kong, China (New York, NY: ACM), 667–677. doi:10.1145/3196709.3196743

Borgdorff, H. (2013). *The Conflict of the Faculties: Perspectives on Artistic Research and Academia*. Amsterdam: Leiden University Press.

Borgdorff, H. (2010). "The Production of Knowledge in Artistic Research," in *The Routledge Companion to Research in the Arts*. Editors M. Biggs and H. Karlsson (New York: Routledge), 44–63. doi:10.4324/9780203841327-12

Braun, V., and Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qual. Res. Psychol.* 3, 77–101. doi:10.1191/1478088706qp063oa

Breazeal, C. L. (2002). *Designing sociable robots*. MIT press.

Briggs, S. R., and Cheek, J. M. (1986). The Role of Factor Analysis in the Development and Evaluation of Personality Scales. *J. Personal.* 54, 106–148. doi:10.1111/j.1467-6494.1986.tb00391.x

Budak, E. P., Zirhli, O., Stokes, A. A., and Akbulut, O. (2016). The Breathing Wall (BRALL)-Triggering Life (In)animate Surfaces. *Leonardo* 49, 162–163. doi:10.1162/leon_a_01199

Cambridge Dictionary, (2021). Soft. Available at: https://dictionary.cambridge.org/dictionary/english/soft (Accessed 25 2.

Carlile, S. (2011). Psychoacoustics. In *The Sonification Handbook, Thomas Hermann, Andy Hunt*, and G. John and Neuhoff (Eds.). Logos Verlag Berlin, Berlin, Germany.

Carpinella, C. M., Wyman, A. B., Perez, M. A., and Stroessner, S. J. (2017). "The Robotic Social Attributes Scale (RoSAS)," in Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction HRI '17, Vienna, Austria, March 2017 (New York, NY: Association for Computing Machinery), 254–262. doi:10.1145/2909824.3020208

Carstensen, L. L., and DeLiema, M. (2018). The positivity effect: A negativity bias in youth fades with age. *Current Opinion in Behavioral Sciences* 19, 7–12. doi:10.1016/j.cobeha.2017.07.009

Cha, E., Kim, Y., Fong, T., and Mataric, M. J. (2018). A survey of nonverbal signaling methods for non-humanoid robots. *Foundations and Trends® in Robotics* 6 (4), 211–323. doi:10.1561/2300000057

Chowning, J., and Bristow, D. (1987). *FM Theory and Applications: By Musicians for Musicians*. Tokyo, Japan: Yamaha Music Foundation.

Collins, Dictionary (2021). Definition of 'soft. https://www.collinsdictionary.com/dictionary/english/soft (Accessed 16, 8 2021).

Cook, P. R. (1999). *Music, cognition, and computerized sound: An introduction to psychoacoustics*. The MIT Press.

Cook, P.R. (2011). "Sound Synthesis for Auditory Display". In (Eds. *Thomas Hermann, Andy Hunt and John G. Neuhoff) The Sonification Handbook*. Berlin, Germany: Logos Publishing House.

Cuan, C., Pakrasi, I., Berl, E., and LaViers, A. (2018a). "CURTAIN and Time to Compile: A Demonstration of an Experimental Testbed for Human-Robot Interaction," in 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), August 2018 (New York, NY: IEEE), 255–261. doi:10.1109/ROMAN.2018.8525520

Cuan, C., Pakrasi, I., and LaViers, A. (2018b). "Perception of Control in Artificial and Human Systems: A Study of Embodied Performance Interactions," in Social Robotics *Lecture Notes in Computer Science*. S. Ge, J.-J. Cabibihan, M. A. Salichs, E. Broadbent, H. He, A. R. Wagner, et al. (Cham: Springer International Publishing), 503–512. doi:10.1007/978-3-030-05204-1_49

Damholdt, M. F., Christina, V., Kryvous, A., Smedegaard, C. V., and Seibt, J. (2019). What Is in Three Words? Exploring a Three-word Methodology for Assessing Impressions of a Social Robot Encounter Online and in Real Life. *J. Behav. Robotics* 10, 438–453. doi:10.1515/pjbr-2019-0034

Della Santina, C., Piazza, C., Gasparri, G. M., Bonilla, M., Catalano, M. G., Grioli, G., et al. (2017). The Quest for Natural Machine Motion: An Open Platform to Fast-Prototyping Articulated Soft Robots. *IEEE Robot. Automat. Mag.* 24, 48–56. doi:10.1109/MRA.2016.2636366

Demers, L.-P. (2014). *Machine Performers: Agents in a Multiple Ontological State*. Plymouth: Plymouth University.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and autonomous systems* 42 (3-4), 177–190. doi:10.1016/s0921-8890(02)00374-3

El-Atab, N., Mishra, R. B., Al-Modaf, F., Joharji, L., Alsharif, A. A., Alamoudi, H., et al. (2020). Soft Actuators for Soft Robotic Applications: A Review. *Adv. Intell. Syst.* 2, 2000128. doi:10.1002/aisy.202000128

Fastl, H., and Zwicker, E. (2007). *Psychoacoustics: facts and models*. Berlin: Springer-Verlag.

Finio, B. (2013). Air-Powered Soft Robotic Gripper. *Instructables*. Available at: https://www.instructables.com/Air-Powered-Soft-Robotic-Gripper/(Accessed July 20, 2021).

Fujisaki, W., Goda, N., Motoyoshi, I., Komatsu, H., and Nishida, S. (2014). Audiovisual integration in the human perception of materials. *Journal of Vision* 14, 12–12. doi:10.1167/14.4.12

Gemeinboeck, P., and Saunders, R. (2019). "Exploring Social Co-presence through Movement in Human-Robot Encounters," in *Movement that Shapes Behaviour: Rethinking How We Can Form Relationships with Non-humanlike, Embodied Agents*. Editors P. Gemeinboeck and E. Jochum, April 2019 (Falmouth, UK: Fallmouth University), 31–27. Available at: http://aisb2019.falmouthgamesacademy.com/proceedings/.

Gemeinboeck, P., and Saunders, R. (2018). "Human-Robot Kinesthetics: Mediating Kinesthetic Experience for Designing Affective Non-humanlike Social Robots," in 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 571–576. doi:10.1109/ROMAN.2018.8525596

Herath, D., McFarlane, J., Jochum, E. A., Grant, J. B., and Tresset, P. (2020). "Arts + Health," in Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction HRI '20, Cambridge, United Kingdom, March 2020 (New York, NY: Association for Computing Machinery), 1–7. doi:10.1145/3371382.3380733

Hu, Y., Zhao, Z., Vimal, A., and Hoffman, G. (2018). "Soft Skin Texture Modulation for Social Robotics," in 2018 IEEE International Conference on Soft Robotics (RoboSoft), Livorno, Italy, March 2018 (New York, NY: IEEE), 182–187. doi:10.1109/ROBOSOFT.2018.8404917

Howard, D.M., and Angus, J.A.S. (2009). *Acoustics and psychoacoustics*. Amsterdam; Boston; London: Focal.

Jee, E. S., Cheong, Y. J., Kim, C. H., Kwon, D. S., and Kobayashi, H. (2009). Sound production for the emotional expression of socially interactive robots, *Advances in human-robot interaction*.

Jee, E. S., Jeong, Y. J., Kim, C. H., and Kobayashi, H. (2010). Sound design for emotion and intention expression of socially interactive robots. *Intelligent Service Robotics* 3 (3), 199–206. doi:10.1007/s11370-010-0070-7

Jekosch, U. (2005). *Assigning meaning to sounds—semiotics in the context of product-sound design Communication acoustics*. Berlin, Heidelberg: Springer, 193–221.

Jenkins, M (2007). *Analog Synthesizers: Understanding, Performing, Buying - From the Legacy of Moog to Software Synthesis*. Focal Press.

Jørgensen, J. (2018). "Appeal and Perceived Naturalness of a Soft Robotic Tentacle," in Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, HRI'18, Chicago, IL, March 2018 (New York, NY: ACM), 139–140. doi:10.1145/3173386.3176985

Jørgensen, J., Bojesen, K. B., and Jochum, E. (2021). Is a Soft Robot More "Natural"? Exploring the Perception of Soft Robotics in Human-Robot Interaction. *Int. J. Soc. Robotics*. doi:10.1007/s12369-021-00761-1

Jørgensen, J. (2019). *Constructing Soft Robot Aesthetics: Art, Sensation, and Materiality in Practice*. Copenhagen: IT University of Copenhagen.

Jørgensen, J. (2017a). "Leveraging Morphological Computation for Expressive Movement Generation in a Soft Robotic Artwork," in Proceedings of the 4th International Conference on Movement Computing MOCO '17, London, United Kingdom, June 2017 (New York, NY: ACM), 1–20. doi:10.1145/3077981.3078029

Jørgensen, J., and Ploetz, S. (2020). "LARPing Human-Robot Interaction," in HRI 2020 Workshop on Exploring Creative Content in Social Robotics, Cambridge, United Kingdom, March 2020 (New York, NY: ACM). Available at: https://portal.findresearcher.sdu.dk/da/publications/larping-human-robot-interaction (Accessed June 19, 2020).

Jørgensen, J. (2017b). "Prolegomena for a Transdisciplinary Investigation into the Materialities of Soft Systems," in ISEA 2017 Manizales: Bio-Creation and Peace: Proceedings of the 23rd International Symposium on Electronic Art, University of Caldas, Manizales, Colombia, June 2017 (Manizales: Department of Visual Design, Universidad de Caldas, and ISEA International), 153–160.

Langeveld, L., van Egmond, R., Jansen, R., and Özcan, E. (2013). Product sound design: Intentional and consequential sounds. *Advances in industrial design engineering* 47 (3).

Laschi, C., Mazzolai, B., and Cianchetti, M. (2016). Soft Robotics: Technologies and Systems Pushing the Boundaries of Robot Abilities. *Sci. Robot.* 1, eaah3690. doi:10.1126/scirobotics.aah3690

LaViers, A., Cuan, C., Maguire, C., Bradley, K., Brooks Mata, K., Nilles, A., et al. (2018). Choreographic and Somatic Approaches for the Development of Expressive Robotic Systems. *Arts* 7, 11. doi:10.3390/arts7020011

Lentz, J.J. (2020). *Psychoacoustics: perception of normal and impaired hearing with audiology applications*. San Diego, CA: Plural Publishing.

Levillain, F., Zibetti, E., and Lefort, S. (2017). Interacting with Non-anthropomorphic Robotic Artworks and Interpreting Their Behaviour. *Int. J. Soc. Robotics* 9, 141–161. doi:10.1007/s12369-016-0381-8

Luo, M., Skorina, E. H., Tao, W., Chen, F., Ozel, S., Sun, Y., et al. (2017). Toward Modular Soft Robotics: Proprioceptive Curvature Sensing and Sliding-Mode Control of Soft Bidirectional Bending Modules. *Soft Robotics* 4, 117–125. doi:10.1089/soro.2016.0041

Merriam Webster (2021). Soft. https://www.merriam-webster.com/dictionary/soft (Accessed 16.8.2021

Milthers, A. D. B., Bjerre Hammer, A., Jung Johansen, J., Jensen, L. G., Jochum, E. A., and Löchtefeld, M. (2019). "The Helpless Soft Robot - Stimulating Human Collaboration through Robotic Movement," in Extended Abstracts of the 2019

CHI Conference on Human Factors in Computing Systems CHI EA '19, Glasgow, United Kingdom, May 2019 (New York, NY: ACM), 1–LBW2421. doi:10.1145/3290607.3312807

Moore, D., Martelaro, N., Ju, W., and Tennent, H. (2017). Making noise intentional: A study of servo sound perception. In 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI),12-21.

Pan, M. K. X. J., Croft, E. A., and Niemeyer, G. (2018). "Evaluating Social Perception of Human-To-Robot Handovers Using the Robot Social Attributes Scale (RoSAS)," in Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, March 2018 (New York, NY: ACM), 443–451. doi:10.1145/3171221.3171257

Read, R., and Belpaeme, T. (2012). "How to Use Non-linguistic Utterances to Convey Emotion in Child-Robot Interaction," in Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction HRI '12, Boston, MA, May 2012 (New York, NY: Association for Computing Machinery), 219–220. doi:10.1145/2157689.2157764

Read, R., and Belpaeme, T. (2014a). "Non-linguistic Utterances Should Be Used Alongside Language, rather Than on Their Own or as a Replacement," in Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction HRI '14, Bielefield, Germany, March 2014 (New York, NY: Association for Computing Machinery), 276–277. doi:10.1145/2559636.2559836

Read, R., and Belpaeme, T. (2013). "People Interpret Robotic Non-linguistic Utterances Categorically," in Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction HRI '13, Tokyo, Japan, March 2013 (New York, NY: IEEE), 209–210. doi:10.1109/hri.2013.6483575

Read, R., and Belpaeme, T. (2014b). "Situational Context Directs How People Affectively Interpret Robotic Non-linguistic Utterances," in Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction HRI '14, Bielefield, Germany, March 2014 (New York, NY: Association for Computing Machinery), 41–48. doi:10.1145/2559636.2559680

Read, R. G., and Belpaeme, T. (2010). "Interpreting Non-linguistic Utterances by Robots," in Proceedings of the 3rd international workshop on Affective interaction in natural environments AFFINE '10, Firenze, Italy, October 2010 (New York, NY: Association for Computing Machinery), 65–70. doi:10.1145/1877826.1877843

Rosenthal-von der Pütten, A. M., and Straßmann, C. (2018). "Interpretation of (In-) Congruent Nonverbal Behavior and Non-linguistic Utterances," in Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction HRI '18, Chicago, IL, March 2018 (New York, NY: Association for Computing Machinery), 221–222. doi:10.1145/3173386.3176964

Rus, D., and Tolley, M. T. (2015). Design, Fabrication and Control of Soft Robots. *Nature* 521, 467–475. doi:10.1038/nature14543

Roland, (2021). The A-to-Z of Synthesizer Terms. Available at: https://rolandcorp.com.au/blog/a-to-z-synthesizer#SGP (Accessed 16, 8 2021).

Shutterly, A., Menguc, Y., and Knight, H. (2018). "Contextual Collision," in Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction HRI '18, Chicago, IL, March 2018 (New York, NY: ACM), 323–324. doi:10.1145/3173386.3176922

Santina, C. D., Piazza, C., Gasparri, G. M., Bonilla, M., Catalano, M. G., Grioli, G., et al. (2017). The Quest for Natural Machine Motion: An Open Platform to Fast-Prototyping Articulated Soft Robots. *IEEE Robotics Automation Magazine* 24, 48–56. doi:10.1109/MRA.2016.2636366

Schwenk, M., and Arras, K. O. (2014). R2-D2 reloaded: a flexible sound synthesis system for sonic human-robot interaction design, The 23rd IEEE International Symposium on Robot and Human Interactive Communication. IEEE, 161–167.

Siedenburg, K., Saitis, C., McAdams, S., Popper, A.N., and Fay, R.R. (2019). Timbre: Acoustics, Perception, and Cognition, *Springer Handbook of Auditory Research Series*. New York, NY: Springer.

Vlachos, E., Jochum, E., and Demers, L.-P. (2018). "HEAT: The Harmony Exoskeleton Self - Assessment Test," in 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Nanjing and Tai'an, China, August 2018 (New York, NY: IEEE), 577–582. doi:10.1109/ROMAN.2018.8525775

Vlachos, E., Jochum, E., and Demers, L.-P. (2016). The Effects of Exposure to Different Social Robots on Attitudes toward Preferences. *Is* 17, 390–404. doi:10.1075/is.17.3.04vla

Walker, J., Zidek, T., Harbel, C., Yoon, S., Strickland, F. S., Kumar, S., et al. (2020). Soft Robotics: A Review of Recent Developments of Pneumatic Soft Actuators. *Actuators* 9, 3. doi:10.3390/act9010003

Wang, L., Nurzaman, S. G., and Iida, F. (2017). Soft-Material Robotics. *FNT in Robotics* 5, 191–259. doi:10.1561/2300000055

Wolfe, H., Peljhan, M., and Visell, Y. (2020). Singing Robots: How Embodiment Affects Emotional Responses to Non-linguistic Utterances. *IEEE Trans. Affective Comput.* 11, 284–295. doi:10.1109/TAFFC.2017.2774815

Zheng, C. Y. (2018). "Affective Touch with Soft Robotic Actuators - A Design Toolkit for Personalised Affective Communication," in Workshop: Reshaping Touch Communication: An Interdisciplinary Research Agenda, ACM CHI Conference on Human Factors in Computing Systems, Montreal, Canada, April 2018 (New York, NY: ACM).

Zheng, C. Y. (2017). "Sentimental Soft Robotics as Companion Artefacts," in 4th International Conference on Movement and Computing (MOCO17), London, United Kingdom, June 2017. Available at: http://moco17.movementcomputing.org/wp-content/uploads/2017/12/ds9-zheng.pdf. (Accessed September 6, 2021).

Zheng, C. Y., and Walker, K. (2019). Soft Grippers Not Only Grasp Fruits: From Affective to Psychotropic HRI, in *2019 Convention of Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB)*. Bath, United Kingdom: AISB.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 05 November 2021
doi: 10.3389/frobt.2021.674292

Modeling and Learning Constraints for Creative Tool Use

Tesca Fitzgerald [1]*, *Ashok Goel* [2] *and Andrea Thomaz* [3]

[1]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, United States, [2]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States, [3]Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX, United States

Improvisation is a hallmark of human creativity and serves a functional purpose in completing everyday tasks with novel resources. This is particularly exhibited in tool-using tasks: When the expected tool for a task is unavailable, humans often are able to replace the expected tool with an atypical one. As robots become more commonplace in human society, we will also expect them to become more skilled at using tools in order to accommodate unexpected variations of tool-using tasks. In order for robots to creatively adapt their use of tools to task variations in a manner similar to humans, they must identify tools that fulfill a set of task constraints that are essential to completing the task successfully yet are initially unknown to the robot. In this paper, we present a high-level process for tool improvisation (tool identification, evaluation, and adaptation), highlight the importance of tooltips in considering tool-task pairings, and describe a method of learning by correction in which the robot learns the constraints from feedback from a human teacher. We demonstrate the efficacy of the learning by correction method for both within-task and across-task transfer on a physical robot.

Keywords: tool manipulation, tool transfer, learning from corrections, human-robot interaction, cognitive robotics

Edited by:
Patricia Alves-Oliveira,
University of Washington,
United States

Reviewed by:
Claude Sammut,
University of New South Wales,
Australia
Paola Ardon,
Heriot-Watt University,
United Kingdom

***Correspondence:**
Tesca Fitzgerald
tesca@cmu.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *01 March 2021*
Accepted: *14 October 2021*
Published: *05 November 2021*

Citation:
Fitzgerald T, Goel A and Thomaz A (2021) Modeling and Learning Constraints for Creative Tool Use. Front. Robot. AI 8:674292. doi: 10.3389/frobt.2021.674292

1 INTRODUCTION

The abundant use of tools for a large range of tasks is a hallmark of human cognition (Vaesen, 2012). Design of new tools for accomplishing novel tasks, as well as improvisation in the absence of typical tools and use of tools in novel ways, are characteristics of human creativity. Consider for example, the design of a paperweight to hold a sheaf of papers, or the use of a paperweight to hammer in a nail if an actual hammer is not available. Both require reasoning about complex relationships that characterizes human cognition and creativity (Penn et al., 2008): The latter task, for instance, requires reasoning about the relationships among the force required to hammer in a nail, the surface of the nail's head, the surface of the paperweight bottom, the weight of the paperweight, and so on.

A robot situated in human society will also encounter environments and tasks suited for human capabilities, and thus it is important for a robot to be able to use human tools for human tasks (Kemp et al., 2007). While a robot may learn to complete a new task with a new tool *via* demonstrations by a human teacher (Argall et al., 2009; Rozo et al., 2013), the demonstration(s) provided for that tool cannot prepare the robot for all variations of that tool it is likely to encounter. These variations can range from different tool dimensions (e.g., different sized spoons, hammers, and screwdrivers) to tool replacements when a typical tool is not available (e.g., using a measuring cup instead of a ladle, or a rock instead of a hammer). An additional challenge is that tools are often used to manipulate other objects in the robot's environment. Given that the shape of a tool alters its effect on its environment

(Sinapov and Stoytchev, 2008), a tool replacement may necessitate a change in the manipulation of that tool in order to achieve the same task goal (Brown and Sammut, 2012).

One aim of developing *creative robots* is to enable robots to exhibit creative reasoning in a similar manner as humans in order to enhance human-robot collaboration. Recently, Gubenko et al. (2021) have called for an interdisciplinary approach that synthesizes conceptual frameworks from diverse disciplines such as psychology, design, and robotics to better understand both human and robot creativity. In human cognition, creative reasoning is exemplified by improvised tool use; particularly, our ability to use analogical reasoning to identify replacement tools or methods that may be used to achieve the original goal, as well as reason over the differences between the original and replacement approaches in order to adapt the replacement to the task (Goel et al., 2020). In design, for example, there is the notion of intrinsic functions and ascribed functions (Houkes and Vermaas, 2010): In the latter, the user can use the object or tool for an ascribed function. Our goals for creative robots are similar: to be able to reason over the suitability of possible tool replacements when the original tool is unavailable, and reason over how the robot's execution of the task must be adapted for the replacement tool.

There are several key challenges in enabling robots to creatively use new tools. First, the robot must **explore** novel tool replacements that support the task constraints. Second, the robot must be able to **evaluate** a novel tool's suitability for a particular task, which involves learning a model of the interactions between the robot's gripper, the tool, objects in the robot's environment that are manipulated by that tool, and how those interactions affect the completion of the task goals. Finally, the robot must **adapt** its task model to the novel tool in order to fulfill these constraints. Prior work has addressed these first two challenges by constructing or identifying creative tool replacements (Choi et al., 2018; Sarathy and Scheutz, 2018; Nair and Chernova, 2020). In this paper, we identify and model the tooltip constraints that play a role in all three of these challenges. In particular, we focus on the third challenge of adapting a robot's task model to a novel tool. The contributions of this paper are as follows:

1) An exploratory analysis of the manipulation constraints that must be fulfilled when using a tool to complete three tasks in simulation.
2) Two models that represent the relationship between the orientation and position constraints when manipulating a tool.
3) An algorithm for training these models using interaction corrections provided by a human teacher, first proposed in Fitzgerald et al. (2019).
4) A discussion of the generalizability of these models when applied to new tools and/or tasks.

We organize the rest of this paper as follows. **Section 2** presents a summary of related work in cognitive science, computational creativity, and robotic tool use. **Section 3** defines the tool transfer problem in terms of constraints on the tooltip pose, which we then explore in **Section 4** *via* an extensive evaluation of the effect of tooltip perturbations on task performance in simulation. In **Section 5**, we discuss how a robot may learn these constraints through corrections provided *via* interaction with a human teacher. Finally, we summarize this paper in **Section 6**.

2 BACKGROUND

2.1 Defining Creative Reasoning

What does it mean for a robot to be "creative"? Prior work in creative robotics has often fallen under one of two categories of creativity: 1) Producing a creative output involving creative domains such as music (Gopinath and Weinberg, 2016) and painting (Schubert and Mombaur, 2013), or 2) Invoking a creative reasoning process. Within the latter category, several criteria for creative reasoning have been proposed, such as autonomy and self-novelty (Bird and Stokes, 2006), in which the robot's creative output is novel to itself but not necessarily to an outside observer. Another definition of a creative reasoning process is one that emphasizes both the variation of potential solutions considered by the agent, as well as the process used to consider and select from those options (Vigorito and Barto, 2008).

Creative reasoning may also be defined in an interactive setting. Co-creativity is a process for creative reasoning in which an agent interacts with a human to iteratively improve upon a shared creative concept. In doing so, co-creativity fosters creative reasoning and may improve the quality of the resulting output (Yannakakis et al., 2014). In prior work, we have defined co-creative reasoning in the context of a robot that collaborates with a human teacher to produce novel motion trajectories, while also aiming to maximize its own, partial-autonomy (Fitzgerald et al., 2017). In the context of a robot reasoning over how it may execute a task in a new environment, this co-creative process allows the robot to obtain the contextual knowledge needed to adapt its task model to meet the constraints of the novel environment.

Creative reasoning has been defined in other relevant domains, such as design creativity. Analogical reasoning is said to be a fundamental process of creativity in design (Goel, 1997). In design by analogy, a new design is created by abstracting and transferring design patterns from a familiar design to a new design problem, where the design patterns may capture relationships among the abstract function, behavior, structure, and geometry of designs. Design also entails discovery of problem constraints (Dym and Brown, 2012) including making implicit constraints in a design problem more explicit (Dabbeeru and Mukerjee, 2011). Fauconnier and Turner (2008) introduced *conceptual blending* as another process for creative reasoning. This approach addresses analogical reasoning and creativity problems by obtaining a creative result from merging two or more concepts to produce a new solution to a problem. Abstraction is enabled by mapping the merged concepts to a *generic space*, which is then grounded in the *blend space* by selecting aspects of either input solution to address each part of the problem. Applied to a robotic agent that uses this creative process to approach a new transfer problem, the robot may combine aspects of several learned tasks to produce a new behavior.

Overall, these methods for creative reasoning highlight two important components of creative reasoning: The exploration of novel solutions to a problem, and an evaluation of each candidate solution's effectiveness. Prior work in creative reasoning (e.g., analogical reasoning, interactive co-creativity, and conceptual blending) have addressed these challenges, but not yet in the context of creative tool use by an embodied robot. This domain requires additional considerations, in that it is grounded in a robot's action and perception (Fitzgerald et al., 2017). First, the robot has imperfect perception of its environment and/or tools, and thus may not have a complete model of the tool(s) it may use. Second, its solution must be in the form of a motion trajectory that utilizes the tool to achieve the task goals. As a result, not only is the *choice* of tool a creative one, but the *usage* of that tool is creative as well. We now review relevant literature that addresses these challenges within the robotic tool use domain.

2.2 Identifying Novel Tool Candidates

Existing work typically focuses on identifying the *affordances* of potential tool candidates. Affordances represent the "action possibilities" that result from the relationship between an object and its environment (Gibson, 1979). Once the affordances of candidate tools have been identified, a robot can reason over the most suitable tool for a particular task and integrate it into its motion plan (Agostini et al., 2015; Choi et al., 2018). However, identifying tool affordances is a non-trivial challenge. Recent work in computer vision has applied deep neural networks to this problem in order to visually predict the affordances for a particular tool (Do et al., 2018). The UMD Part Affordance Dataset (Myers et al., 2015) is intended to support further work on visual affordance detection. This dataset contains RGB-D images for 105 tools, grouped into 17 object categories. Each tool is photographed at roughly 75 orientations, each of which corresponds to a pixel-wise labeling according to 7 possible affordances (e.g., cutting, grasping, pounding). Other, physics-based features such as the dimensions or material of an object may also be used to judge their effectiveness as potential tools, such as when identifying a pipe as a makeshift lever to pry open a door (Levihn and Stilman, 2014). Prior work has shown that, in addition to using demonstrations to learn a task, a robot may also use demonstrations to learn to recognize the affordance-bearing subparts of a tool such that it can identify them on novel objects (Kroemer et al., 2012).

When a suitable tool replacement is not already available in the robot's environment, it may be necessary to assemble one (Sarathy and Scheutz, 2018). Choi et al. (2018) extends the ICARUS cognitive architecture to assemble virtual tools from blocks. Nair et al. (2019) describes a method for tool construction by pairing candidate tool parts and then evaluating each pair by the suitability of the shape and attachability of the two parts. Later work (Nair and Chernova, 2020) integrates this process into a planning framework such that the task plan includes both the construction and use of the required tool.

While candidate tool identification is not the focus of this article, it is an essential step in our eventual goal of creative tool use. Overall, prior work on this topic demonstrates the task-specific requirements for identifying novel tool candidates, and the importance of identifying the salient features of a tool within the context of the current task. We now consider how these features affect the tool's suitability when evaluating them for a particular task.

2.3 Evaluating Novel Tool Candidates

The shape of a tool alters its effect on its environment (Sinapov and Stoytchev, 2008), and thus a tool replacement may necessitate a change in the manipulation of that tool in order to achieve the same task goal (Brown and Sammut, 2012). For tasks involving the use of a rigid tool, the static relationship between the robot's hand and the tooltip is sufficient for controlling the tool to complete a task (Kemp and Edsinger, 2006; Hoffmann et al., 2014). These methods assume a single tooltip for each tool, and that this tooltip is detected *via* visual or tactile means. For tasks involving multiple surfaces of the tool, the task model can be explicitly defined with respect to those segments of the tool, and repeated with tools consisting of similar segments (Gajewski et al., 2018). However, this assumes a hand-defined model that represents the task with respect to pre-defined object segments, and that these object segments are shared across tools. Given enough training examples of a task, a robot can learn a success classifier that can later be used to self-supervise learning task-oriented tool grasps and manipulation policies for unseen tools (Fang et al., 2018). We similarly aim to situate a new tool in the context of a known task, but eliminate the assumptions that 1) the new tool is within the scope of the training examples (which would exclude creative tool replacements) and 2) that the tool features relevant to the task are observable and recorded by the robot.

2.4 Adapting Task Models to Novel Tools

The aim of transfer learning for reinforcement learning domains is typically to use feedback obtained during exploration of a new environment in order to enable reuse of a previously learned model (Taylor and Stone, 2009). In previous work, we have shown how interaction can be used to transfer the high-level ordering of task steps to a series of new objects in a target domain (Fitzgerald et al., 2018). Similarly, the aim of one-shot learning is to quickly learn a new task, often improving learning from a single demonstration by adapting previous task knowledge. Prior work in this space focuses on learning a latent space for the task in order to account for new robot dynamics (Srinivas et al., 2018) or new task dynamics (Fu et al., 2016; Killian et al., 2017). "Meta-learning" approaches have succeeded at reusing visuomotor task policies learned from one demonstration (Chelsea et al., 2017) and using a new goal state to condition a learned task network such that it can be reused with additional task objects (Duan et al., 2017). We address the problem of a robot that has *not yet* been able to explore these relationships, aiming to enable rapid adaptation of a task model for unseen task/parameter

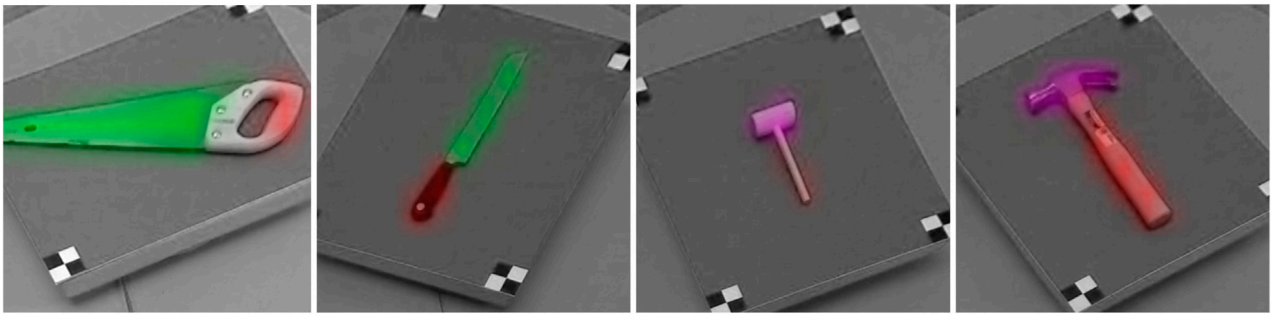


FIGURE 1 | Affordance regions may be broad, spanning multiple possible tooltips. As a result, predicting the affordance region is not sufficient to plan with respect to that tool's tooltip. For example, the full blade surfaces of the saw and knife are labeled as enabling the "cutting" affordance (highlighted in green) and the "grasping" affordance (highlighted in red); however, cutting is only performed using the edge of the blade, and requires that the blade be oriented toward the cutting target. Similarly, different points of a hammer head may enable different tasks (e.g., pounding versus prying), and thus detecting a task-independent affordance region (highlighted in purple) is not sufficient to plan a task trajectory.

relationships. The tool transform models learned by our approach are not specific to any task learning algorithm or representation, and thus can compliment or bootstrap methods for reinforcement, one-shot, and meta learning.

2.5 Summary of Related Work

Through prior work, we have identified three key steps for creative tool use: Exploring novel tools, evaluating novel tools, and adapting task models to novel tools. These stages are not entirely separable from each other, as evaluating reflects how well the robot anticipates being able to adapt its task model for a particular tool, and exploration results in a set of tools that meet some criteria such that they may be evaluated in the context of the task. A common theme through all three steps is the importance of *constraints* (e.g., tool shape, segments, or visual features) that dictate how a task model may be adapted to a particular tool, and as a result, play a role in the exploration and evaluation steps as well.

In the rest of this paper, we focus on this challenge of identifying and modeling *constraints*, and demonstrate how these constraints may be used in the evaluating and adapting steps of creative tool use. While we do not explicitly address creative tool exploration, we aim for this work to support future research on identifying these constraints visually to enable this exploration.

3 TOOLTIPS AS CONSTRAINTS

Suppose that a robot has learned a trajectory $\mathbf{T_a} = [p_a^{(0)}, p_a^{(1)}, \ldots, p_a^{(n)}]$ consisting of end-effector poses $p_a^{(i)}$ for a particular task using tool a, and now must complete the same task using a different tool b. Our goal is to transform each pose individually for tool b. Representing an original pose for tool a in terms of its 3×1 translational vector $\mathbf{t_a}$ and 4×1 rotational vector $\mathbf{r_a}$, we transform it into a pose $\mathbf{p_b}$ for tool b as follows:

$$\mathbf{p_b} = \phi_a^b(\mathbf{p_a}) = \langle \mathbf{t_a} + \hat{\mathbf{t}}, \mathbf{r_a} \cdot \hat{\mathbf{r}} \rangle \tag{1}$$

Here, $\mathbf{r_a} \cdot \hat{\mathbf{r}}$ refers to the Hamilton product between the two quaternions. This definition relies on a known transform between tools a and b, which requires knowledge of the appropriate "reference" point for both tools such that their transform can be computed. Neither reference point is initially known by the robot, however, nor can it be extracted from the trajectory which is represented according to the robot's end-effector, and not according to any point on the tool itself.

Identifying the "reference point" for a tool is non-trivial. While prior work has addressed the problem of identifying affordance regions of a tool, these regions are too broad to characterize the transform between two tools. **Figure 1** illustrates examples of these labeled affordance regions based on the UMD Part Affordance Dataset (Myers et al., 2015). While this dataset is relevant to identifying similar regions on two separate tools, it does not address the problem of specifying the equivalent points of a tool that may be used to transform the trajectory for a *particular task* from one tool to another. For example, the full blade of a knife may be labeled as enabling the "cutting" affordance (**Figure 1**), even though a cutting task is likely to be performed with respect to only the edge of the blade. Furthermore, since affordance data is presented in the form of pixel-wise image labels, it does not provide any data concerning the kinematic implications of using this tool. Since the tool is observed and labeled from a static, overhead perspective, affordance data is only available along a single 2D plane, and thus does not indicate the *orientation* at which each affordance is or is not valid.

This is essential for manipulating the tool properly; even if the robot were to determine that the relevant surface of a knife is located along the edge of its blade, the blade must still be oriented carefully with respect to the cutting target for the task to be completed successfully. We refer to the acting surface of the tool (e.g., a singular point along the edge of the knife blade, or a singular point on a mallet's pounding surface) as a **tooltip** that is defined by a pose containing both the position and orientation of that tooltip. In summary, we expect that **successful task completion relies on the robot having a model of the composite transform between 1) the end-effector, 2) its grasp of the tool** (highlighted in red in **Figure 1**), **and 3) the tooltip position and orientation**.

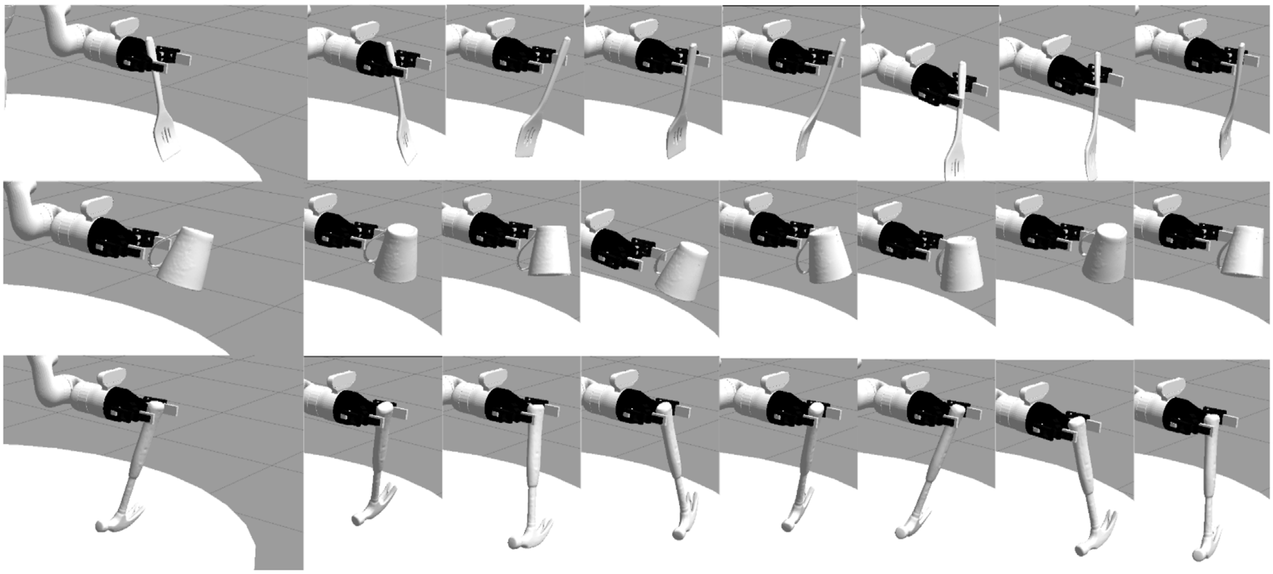


FIGURE 2 | We performed an evaluation across three tools: a spatula, mug, and hammer. For each tool, we perturbed the trajectory of the tooltip by adjusting the robot's grasp of the tool. These pose variations are just a small set of the 729 perturbations we evaluated for each tool-task pairing.

While we may mathematically represent a tooltip as a singular pose, practically, however, there are likely many possible tooltips that may lead to successful task execution. Additionally, the constraint over the tooltip may also differ depending on the context in which it is used: The orientation of a hammer is constrained along two axes when hammering a nail, but the hammer may still be rotated around the nail (e.g., its "yaw" rotation) without affecting task performance. This example supports the notion of a *one-to-many* relationship between 1) a tooltip and 2) the tool poses that enable that tooltip to be used.

In the remainder of this paper, we explore this one-to-many relationship. In **Section 4**, we demonstrate how a single tooltip can be expanded into a set of effective tool poses, thus highlighting the challenges of learning tooltip constraints. In **Section 5**, we consider this relationship in the opposite direction, and present two models for deriving a single tooltip from a set of valid poses demonstrated by a human teacher.

4 CHARACTERIZING TOOL CONSTRAINTS

We first explore the effect of tooltip constraints by expanding a single tooltip into a set of tool poses that result in successful task execution. To do so, we transform a trajectory that results in successful task execution (and thus the tooltip is implicitly-defined) such that the tooltip's trajectory is perturbed slightly. In doing so, we can evaluate the effect of that perturbation on task performance, and ultimately model the constraints that dictate which poses result in successful use of the tooltip.

In this section, we address two key research questions:

1) How do changes in tool pose affect task performance?
2) How do the constraints on tool pose differ across tools and/or tasks?

4.1 Evaluating Tool-Task Constraints in Simulation

We address these research questions by evaluating the performance of a large set of trajectory perturbations using a simulated 7-DOF Kinova Gen3 robot arm situated on a round table in a Gazebo simulated environment. We evaluated the effect of trajectory perturbations on three tools: A hammer, a mug, and a spatula (**Figure 2**). We fixed the robot's grasp as a static transform between the robot's gripper and the tool, and thus did not evaluate the effects of the robot's grasp strength or stability on tool use.

For each tool, we provided a demonstration of three tasks: Hooking (**Figure 3A**), lifting (**Figure 3B**), and sweeping (**Figure 3C**). Each demonstration was provided in a Gazebo simulator as a set of end-effector keyframes. Depending on the tool being demonstrated, this resulted in 5-7 keyframes for hooking, 4-6 for lifting, and 13-18 for sweeping. These end-effector keyframes were then converted to keyframe trajectories represented in the robot's joint-space. We used the MoveIt (Coleman et al., 2014) implementation of the RRTConnect planner to plan between joint poses during trajectory execution. We simulated a trajectory perturbation by altering the rigid transform between the robot's gripper and the tool itself, according to a pre-determined set of position and orientation alterations that are consistent across all tools and tasks. As a result, each trajectory perturbation is identical with respect to the robot's end-effector, but differ with respect to the trajectory of the tool itself. This allowed us to use the same joint-space trajectory

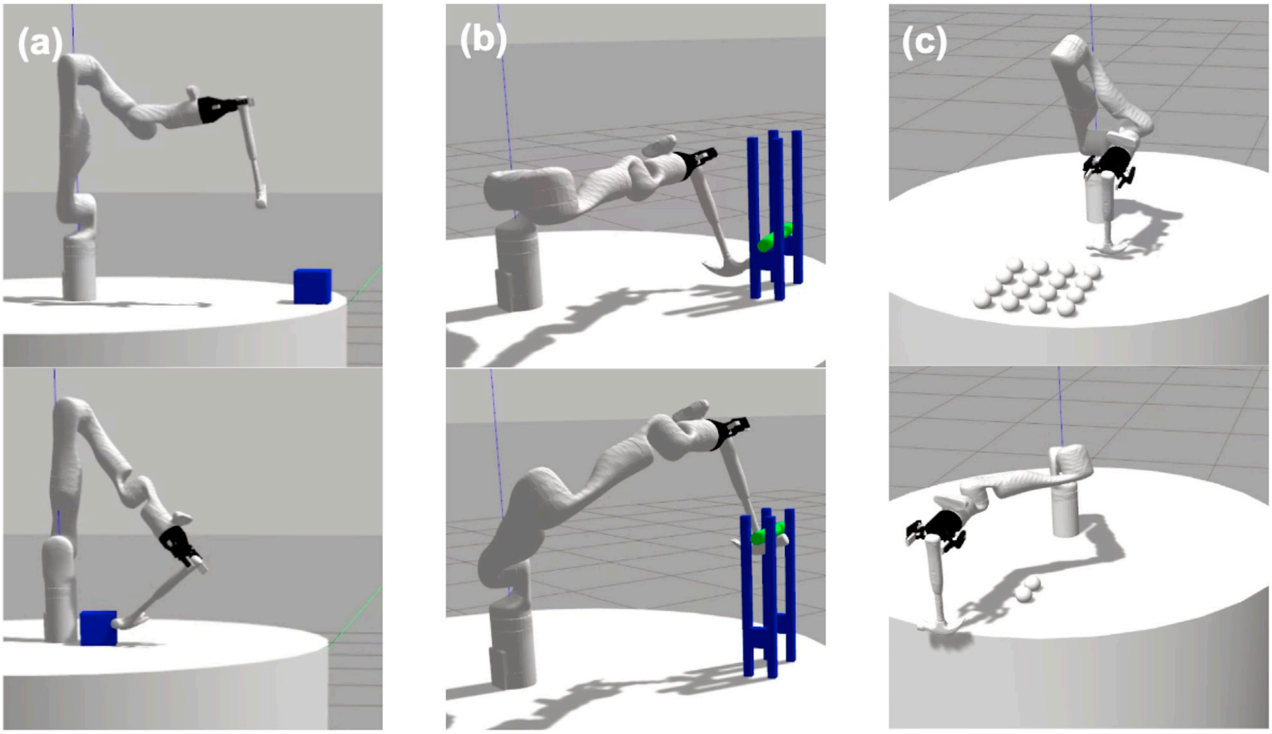


FIGURE 3 | Initial and goal states for the **(A)** hooking, **(B)** lifting, and **(C)** sweeping tasks.

for all perturbations of a single tool-task pairing, thus reducing the likelihood of planning errors across all perturbations and also minimizing any changes in the robot's joint motion that might affect task performance. Despite the same trajectory being executed across all perturbations of a single tool-task pairing, planning errors may still occur when a perturbation results in the tool colliding with its environment, thus preventing the rest of the trajectory from being executed.

Each perturbation resulted from a unique permutation of changes applied to the tool's demonstrated position along the x, y, and z axes and demonstrated orientation along the roll, pitch, and yaw axes. The tool's x, y, and z positions were each configured at one of three distances from the demonstrated tool position: $[-0.01, 0, 0.01]$ meters. The tool's roll, pitch, and yaw rotations were each configured at one of three angles from the demonstrated tool orientation: $[-\frac{\pi}{16}, 0, \frac{\pi}{16}]$ radians. These position and orientation perturbations were empirically chosen such that, when combined, their effect on task performance can be observed on a spectrum. We observed that larger ranges of pose or orientation changes would be less likely to result in completion of any aspect of the task, whereas smaller ranges may not fully explore the range of successful perturbations. However, as we note later in **Section 4.3**, we observe that different tools vary in their sensitivity to these perturbations, and thus a more fine-grained set of perturbations should be explored in future work.

Overall, the permutation of these configurations resulted in a total of $3^6 = 729$ perturbations for each tool-task pairing. We executed each perturbation twice in simulation (to account for the non-deterministic effects of the simulator dynamics) and recorded the average performance of the two trials, with performance being measured according to task-specific measures. All performance metrics were scaled to a 0–1 range. In the hooking task, performance was measured as the distance (in meters) between the box and the robot's base, with less distance correlating to higher performance. The initial and goal states of this task are shown in **Figure 3A**. In the lifting task, the robot's performance was measured as the green bar's height above the table (in meters). A small number of trials resulted in the bar being removed from the support structure entirely. In these cases, we recorded the performance as that of the task's initial state (i.e., a failure case). **Figure 3B** shows the initial and goal states of this task. In the sweeping task, performance is measured as the number of spheres that were swept off the table, with maximum performance being 16 spheres. The initial and goal states of this task are shown in **Figure 3C**.

4.2 Results

Our evaluation measured how sensitive each tool-task pairing is to perturbations of the tooltip's trajectory: The more sensitive the tool-task pairing is to perturbations, the more likely that a perturbation will lead to a task failure. Low task performance may be caused by the tooltip no longer contacting any relevant objects in the task (and thus leaving the task in its initial state), or by collisions between the tool's new configuration and its environment that prevent the robot from executing the full

(A)

Tasks	Mug	Spatula	Hammer
Sweeping	3.29%	35.11%	10.01%
Hooking	3.29%	35.12%	9.19%
Lifting	4.25%	35.80%	45.27%

(B)

Tasks	Mug	Spatula	Hammer
Sweeping	$\mu = 0.77$ $\sigma = 0.13$	$\mu = 0.82$ $\sigma = 0.14$	$\mu = 0.58$ $\sigma = 0.21$
Hooking	$\mu = 0.85$ $\sigma = 0.19$	$\mu = 0.73$ $\sigma = 0.27$	$\mu = 0.63$ $\sigma = 0.30$
Lifting	$\mu = 0.83$ $\sigma = 0.18$	$\mu = 0.70$ $\sigma = 0.23$	$\mu = 0.70$ $\sigma = 0.18$

FIGURE 4 | (A) Percentage of failed trials (performance ≤ 0.05). Darker cells indicate higher percentage of failed trials. **(B)** Mean and standard deviation performance of thresholded (performance > 0.05) trials. Darker cells indicate higher mean performance.

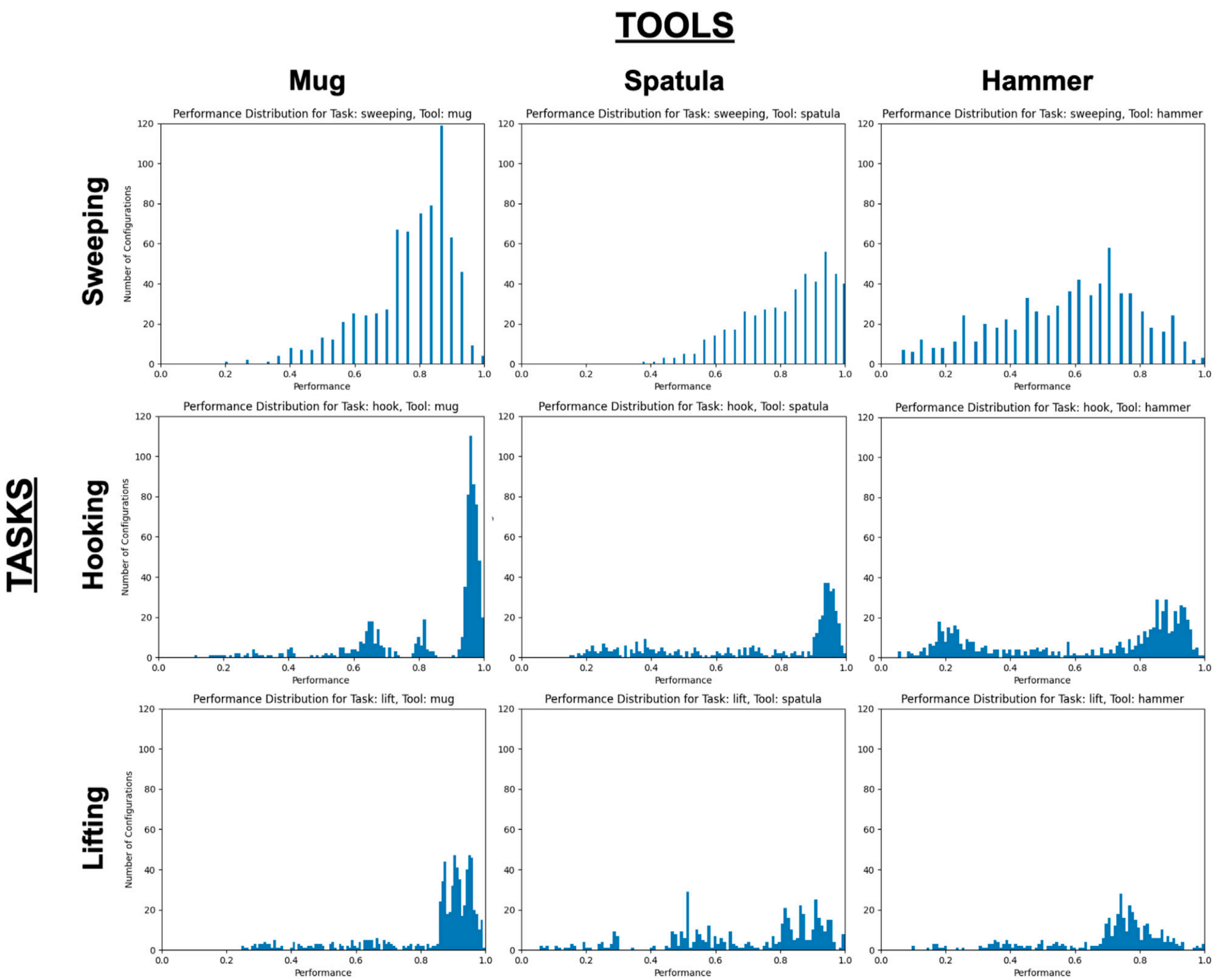


FIGURE 5 | Performance distributions over all tool-task pairings, with all trials with performance ≤ 0.05 excluded. *X*- and *Y*-axes are consistent across all graphs.

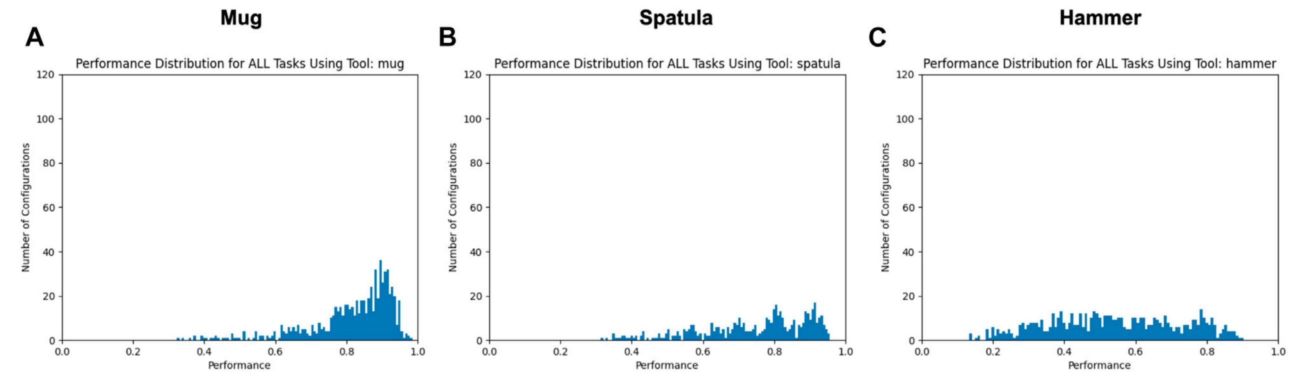


FIGURE 6 | Mean performance distributions using each tool for all tasks, with all trials with mean performance ≤ 0.05 excluded. *X*- and *Y*-axes are consistent across all graphs.

trajectory. We set a threshold performance of 0.05 (on a 0–1 scale), and report the percentage of perturbations that fail to exceed this threshold in **Figure 4**.

We include only the set of perturbations that exceed this threshold in the histograms in **Figure 5**, which illustrate the performance distributions over the set of perturbations exceeding this threshold. Since the original, unperturbed pose is already known to achieve near-optimal task performance, these graphs illustrate how many perturbations of that original pose still fulfill the tooltip constraints and result in high performance (i.e., the perturbations resulting in the peak observed near $x = 1.0$ on each graph). We report the mean and variance over these performance results in (**Figure 4B**).

Figure 6 shows the distribution over the *mean performance* over all three tasks; that is, the performance metric for each perturbation is the average of its performance on the sweeping, hooking, and lifting tasks. We again only consider datapoints above a performance threshold > 0.05 in order to focus on the set of *valid* tooltip constraints for each tool.

4.3 Discussion

Research Question #1: How do changes in tool pose affect task performance? **The relationship between performance and tool pose may be non-linear.** If this relationship were linear, we would expect **Figure 5** to primarily contain Gaussian-like performance distributions, such that as the robot evaluates trajectory perturbations further from the original trajectory, its performance resulting from those perturbations decreases proportionally. While this is the case in some tool-task pairings (e.g., all tools used for the sweeping task, and the lifting task using the hammer), other performance distributions appear to be bi-modal in nature (e.g., using the hammer in the hooking task or using the spatula for lifting) or contain several peaks (e.g., using the mug for hooking). This suggests that there is a non-linear relationship between changes in the tool pose, and its resulting effects on task performance. Note that in our evaluation, we applied trajectory perturbations according to the single tooltip that was demonstrated for each tool-task pairing. An opportunity for future research is the identification of alternate tooltips based on the tool's shape or structure.

Research Question #2: How do the constraints on tool pose differ across tools and/or tasks? **Tools differed in their sensitivity to pose changes.** For example, using the spatula tool resulted in the highest percentage of failed trials (35.11–35.8%) across all three tasks, while the mug resulted in the lowest (3.29–4.25%) across all three tasks. One hypothesis for this performance difference is that since the mug was the smallest tool, changes in the tool pose had a smaller effect on its *tooltip* pose in comparison to the taller tools (spatula and hammer). We observed widely varying failure rates when using the hammer, ranging from 9.19 to 10.01% on the hooking and sweeping tasks, respectively, and 45.27% on the lifting task. One reason for this performance difference may be that a different tooltip was used for the lifting task compared to the hooking and sweeping tasks. In the former, the robot uses a "corner" of the hammer to lift the bar (**Figure 3B**), whereas the hooking and sweeping tasks use a wider surface area of the hammer as a tooltip. This may provide more tolerance to pose perturbations. **Overall, this suggests that the sensitivity of tooltip constraints depends on the surface of the tool being used.**

Figure 6 also supports this hypothesis. These distribution graphs reflect the consistency in tooltip constraints across tasks. While the geometry of the tool itself remains constant across tasks, the same tooltip is not necessarily used across tasks (e.g., using separate surfaces of the hammer for sweeping vs lifting). The reduced performance shown in these graphs (in comparison to **Figure 5**) indicates that **the tooltip constraints applied to one task may not be generalizable to other tasks using the same tool.**

We now consider the challenge of how a robot may quickly learn these constraints in the context of a new tool, and whether we can model the instances in which a robot can reuse a learned tooltip model in the context of another task. While a robot can learn to use a tool through demonstrations, the one-to-many mapping between tooltip constraints and the set of tool poses that meet those constraints means that there are many possible demonstrations that a robot may receive for a tool/task pairing. Learning the underlying tool constraint is therefore a challenge, as the teacher is providing demonstrations that sample from an unknown, underlying relationship between the end-effector and the tooltip. In the next section, we explore how a robot can utilize corrections in order to model and learn the underlying tooltip constraint.

5 LEARNING CONSTRAINTS FROM INTERACTIVE CORRECTIONS

In the previous section, we evaluated the one-to-many mapping between tooltips constraints and end-effector poses that meet those constraints. In order to adapt the robot's task model to a novel tool, however, we also need to analyze this mapping in the reverse direction: inferring the underlying tooltip constraint that has resulted in a set of corresponding end-effector poses.

We address this challenge in the context of a robot that learns from demonstrations by a human teacher who is familiar with the task and tool that the robot aims to use. By comparing two trajectories, each using a separate tool to complete the same task, we aim to model the relationship between the two tooltips constraints such that it can be reused in the context of another task.

While a robot can quickly receive demonstrations (Argall et al., 2009; Chernova and Thomaz, 2014) using a new tool, these demonstrations may not be sufficient to learn the underlying tooltip constraints. Due to the unstructured nature of task demonstrations, the two demonstrations (each provided using a different tool) may vary in ways that do not reflect how the task should be adapted based on which tool is used. For example, the teacher may choose a different strategy for completing the task with the second tool, or the robot may be starting from a new arm configuration when the teacher demonstrates the task with the second tool. For these reasons, we utilize *corrections* of the robot's behavior, which have been shown to be effective interface for adapting a previously-learned task model (Argall et al., 2010; Sauser et al., 2012; Bajcsy et al., 2018). Rather than have the teacher provide a new demonstration using the new tool, the robot attempts to complete the task on its own and is interrupted and corrected by the teacher throughout its motion. As a result, this interaction results in a series of correction pairs, where each pair represents the robot's originally-intended end-effector pose and its corresponding, corrected pose that was indicated by the teacher.

Our research questions are as follows:

1) How can we model a tooltip constraint using data provided *via* sparse, noisy corrections?
2) Under what conditions can the tooltip constraints learned from corrections on one task be used to adapt other task models to the same replacement tool? What characteristics of the tool and task predict whether a previously-learned tooltip constraint can be applied?

In the following sections, we address these research questions using the Transfer by Correction algorithm, which we first described in Fitzgerald et al. (2019).

5.1 Problem Definition

We assume that each demonstration consists of a series of keyframes (Akgun et al., 2012). The robot receives corrections by executing a trajectory planned using the original task model, pausing after a time interval defined by the keyframe timings set during the original demonstration. The teacher then moves the robot's gripper to the correct position, after which the robot resumes task execution for the next time interval, repeating the correction process until the entire task is complete. Each resulting correction at interval i consists of the original pose $\mathbf{C_a^i}$ (using tool a) and the corrected pose $\mathbf{C_b^i}$ (using new tool b) at keyframe i. A collection of K corrections (one for each of K keyframes) results in a K x 2 correction matrix:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C_a^0} & \mathbf{C_b^0} \\ \mathbf{C_a^1} & \mathbf{C_b^1} \\ \dots & \\ \mathbf{C_a^K} & \mathbf{C_b^K} \end{bmatrix} \tag{2}$$

Each corrected pose $\mathbf{C_b^i}$ provides a sample of the transfer function value with the original pose $\mathbf{C_a^i}$ at keyframe i as input, plus some amount of error from the optimal correction pose:

$$\mathbf{C_b^i} = \phi_a^b(\mathbf{C_a^i}) + \epsilon \qquad \epsilon_n \sim \mathcal{N}(0, \sigma_n^2) \tag{3}$$

We assume ϵ is sampled from a Gaussian noise model for each axis $n \in [1\ldots6]$ of the 6D end-effector pose. Our aim is to learn a transfer function ϕ that optimally reflects the tooltip constraints, using a correction matrix $\mathbf{C}$.

5.2 Approach: Transfer by Correction

Given a task trajectory $\mathbf{T}$ for tool a consisting of a series of t poses in task space such that $\mathbf{T} = [\mathbf{p_0}, \mathbf{p_1}, \ldots, \mathbf{p_t}]$, we transform each pose individually for tool b. Representing an original pose for tool a in terms of its 3×1 translational vector $\mathbf{t_a}$ and 4×1 rotational vector $\mathbf{r_a}$, we transform it into a pose $\mathbf{p_b}$ for tool b as follows:

$$\mathbf{p_b} = \phi_a^b(\mathbf{p_a}) = \langle \mathbf{t_a} + \hat{\mathbf{t}}, \mathbf{r_a} \cdot \hat{\mathbf{r}} \rangle \tag{4}$$

Here, $\mathbf{r_a} \cdot \hat{\mathbf{r}}$ refers to the Hamilton product between the two quaternions. The goal is now to estimate the optimal rotational $\hat{\mathbf{r}}$ and translational $\hat{\mathbf{t}}$ transformation components from the corrections matrix $\mathbf{C}$, and then apply these transformations to the trajectory $\mathbf{T}$. Our approach addresses this goal by **(1)** modeling $\mathbf{C}$, particularly the relationship between each correction's translational and rotational components, 2) sampling a typical translational transformation $\hat{\mathbf{t}}$ and rotational transformation $\hat{\mathbf{r}}$ from this transform model, and 3) applying $\hat{\mathbf{t}}$ and $\hat{\mathbf{r}}$ to transform each pose in the task trajectory according to **Equation 4**.

5.3 Task Constraints

We observe that corrections indicate constraints of the tooltip's position and/or orientation, and that these constraints are reflected in the relationship between the translation and rotation components of each correction. Broadly, each correction may primarily indicate:

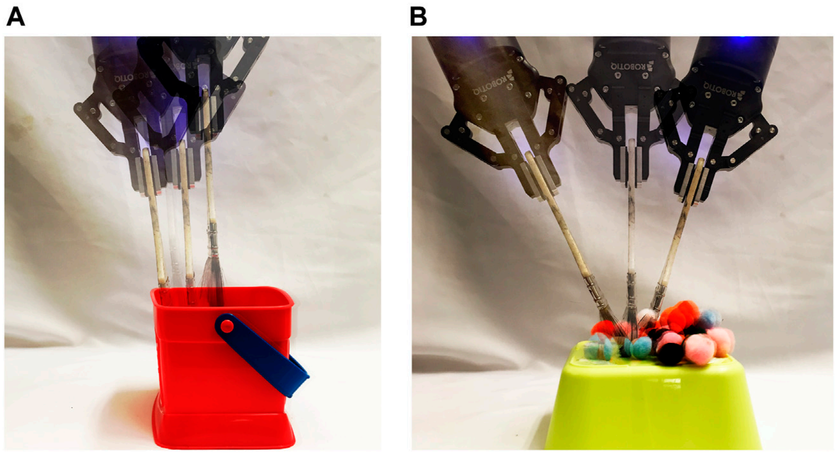


FIGURE 7 | Poses meeting the same orientation constraint share similar orientations but vary more in their position **(A)**, whereas poses meeting the same center-of-rotation constraint rotate around the tooltip **(B)**.

- An *unconstrained* point in the trajectory, and thus should be omitted from the tool transform model.
- An *orientation constraint*, where the rotation of the tooltip (and thus the end effector) is constrained more than its position (e.g., hooking a box is constrained more by the orientation of the hook than its position, as in the left of **Figure 7**).
- A *center-of-rotation constraint*, where the position of the tooltip is constrained more than its rotation (e.g., sweeping a surface with a brush). Note that the *tooltip* position is the center of this constraint rather than the end-effector itself, and thus the range of valid end-effector poses forms an arc around the tooltip, and its orientation remains angled toward the tooltip (e.g., **Figure 7B**).

We define two *tool transform models*, first presented in Fitzgerald et al. (2019), each reflecting either orientation or center-of-rotation constraints. We fit the corrections matrix to each tool transform model, using RANSAC (Fischler and Bolles, 1981) to iteratively estimate the parameters of each model while discarding outlier and unconstrained correction data points. Each iteration involves 1) Fitting parameter values to a sample of n datapoints, 2) Identifying a set of inlier points that also fit those model parameters within an error bound of ϵ, and 3) Storing the parameter values if the inlier set represents a ratio of the dataset > d. The RANSAC algorithm relies on a method for fitting parameters to the sample data, and a distance metric for a datapoint based on the model parameters. These are not defined by the RANSAC algorithm, and so we specify the parameterization and distance metric according to the tool transform model used, which we describe more in the following sections. We define an additional method to convert the best-fitting parameters following RANSAC completion into a typical transform that can be applied to poses.

5.4 Linear Tool Transform Model

Based on the *orientation* constraint type, we first consider a linear model for correction data, where corrections fitting this model share a linear relationship between the translational components of the corrections, while maintaining a constant relationship between the rotational components of corrections (**Figure 8A**). We model this linear relationship as a series of coefficients obtained by applying PCA to reduce the 3D position corrections to a 1D space.

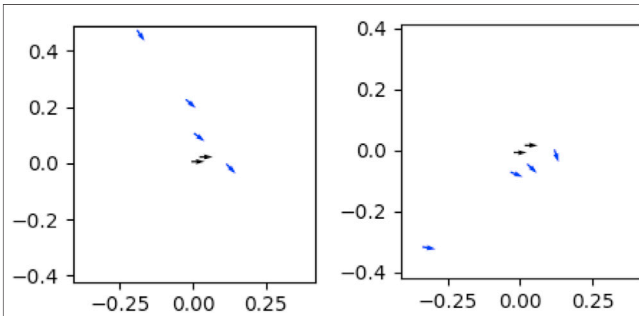


FIGURE 8 | Each plot represents one set of corrections for a task. The position of each arrow represents the change in < x, y > position, and points in the direction of the change in orientation introduced by that correction. Orientation constraints can be seen in **(A)**, where the majority of corrections on this tool have low variance in their orientation, but higher variance in their *x-y* position. Center-of-rotation constraints can be seen in **(B)**, where the majority of corrections arc around a singular center of rotation, and orientation is dependent on the *x-y* position. Unconstrained keyframes (colored grey) are located near (0,0).

5.4.1 RANSAC Algorithm Parameters

The RANSAC algorithm is performed for k iterations, where we use the estimation

$$k = \frac{log(1.0 - p)}{log(1.0 - w^n))} \tag{5}$$

with desired confidence $p = 0.99$ and estimated inlier ratio $w = 0.5$. Additional parameters are as follows: $n = 2$ is the number of data points sampled at each RANSAC iteration, $\epsilon = 0.01$ is the error threshold used to determine whether a data point fits the model, and $d = 0.5$ is the minimum ratio between inlier and outlier data points in order for the model to be retained.

5.4.2 Model Parameter Fitting

Model fitting during each iteration of RANSAC consists of reducing the datapoints to a 1D model using PCA, returning the mean translational correction and the coefficients for the first principal component of the sample $\mathbf{S}$:

$$\Theta_{\text{linear}}(\mathbf{S}) = \langle \theta_\mu, \theta_\mathbf{u} \rangle \qquad \theta_\mu = \frac{1}{|\mathbf{S}|}\sum_{\mathbf{p}\in\mathbf{S}} \mathbf{p_t} \tag{6}$$

where $\mathbf{p_t}$ is the 3×1 translational difference indicated by the correction $\mathbf{p}$, $\mathbf{S}$ is the subset of the corrections matrix $\mathbf{C}$ sampled during one iteration of RANSAC such that $\mathbf{S} \subset \mathbf{C}$, and $\theta_\mathbf{u}$ is the eigenvector corresponding to the largest eigenvalue of the covariance matrix $\Sigma = \frac{1}{|\mathbf{S}|}\mathbf{S_t}^T\mathbf{S_t}$.

5.4.3 Error Function

Each iteration of RANSAC calculates the total error over all data points fitting that iteration's model parameters. We define the error of a single correction datapoint $\mathbf{p}$ as the sum of its reconstruction error and difference from the average orientation correction, given the current model parameters θ:

$$\delta_{\text{linear}}(\mathbf{p}, \theta) = \left\|\mathbf{p_t} - \left(\theta_\mu + \left(\mathbf{p_t} - \theta_\mu\right)^T \theta_\mathbf{u}\theta_\mathbf{u}^{\mathbf{T}+}\right)\right\| + \gamma\left(\left(1 - \bar{\mathbf{q}}_\mathbf{n}\mathbf{p}_\mathbf{n}^T\right)^2\right) \tag{7}$$

where $\mathbf{x}^+$ indicates the Moore-Penrose pseudo-inverse of a vector, $\mathbf{p_n}$ is the unit vector representing the orientation difference indicated by the correction $\mathbf{p}$, $\bar{\mathbf{q}}_\mathbf{n}$ is a unit vector in the direction of the average rotation sampled from the model (defined in the next section), and γ is the weight assigned to rotational error ($\gamma = 1$ in our evaluations).

5.4.4 Sampling Function

After RANSAC returns the optimal model parameters and corresponding set of inlier points $\hat{\mathbf{I}} \subset \mathbf{C}$, the rotation and translation components of the transformation are sampled from the model. We define the sampling function according to the estimated "average" rotation $\bar{\mathbf{q}}$:

$$\Psi(\hat{\mathbf{I}}, \hat{\theta})_{\text{linear}} = \langle \bar{\mathbf{q}}, \bar{\mathbf{t}} \rangle \quad \bar{\mathbf{q}} = \underset{\mathbf{q}\in\mathbb{S}^3}{\arg\max}\, \mathbf{q}^T M \mathbf{q} \quad M = \frac{1}{|\hat{\mathbf{I}}|}\sum_{\mathbf{p}\in\hat{\mathbf{I}}} \mathbf{p}_\mathbf{q}^i \mathbf{p}_\mathbf{q}^{i\,T} \tag{8}$$

The solution to $\bar{\mathbf{q}}$ for this maximization problem is the eigenvector corresponding to the largest eigenvalue of M (Markley et al., 2007). The sample translation $\bar{\mathbf{t}}$ is the 3D offset corresponding to the mean value $\bar{z}$ from the 1D projection space:

$$\bar{\mathbf{t}} = \hat{\theta}_\mu + \bar{z}\hat{\theta}_\mathbf{u}^{T+} \qquad \bar{z} = \frac{1}{|\hat{\mathbf{I}}|}\sum_{p\in\hat{\mathbf{I}}}\left(\mathbf{p}_t - \hat{\theta}_\mu\right)^T\hat{\theta}_\mathbf{u} \tag{9}$$

5.5 Rotational Tool Transform Model

We now consider a model for corrections reflecting a *center-of-rotation constraint*, in which we make the assumption that corrections indicate a constraint over the tool tip's *position*. Since the tool tip is offset from the end-effector, the position and rotation of the end-effector are constrained by each other such that the end-effector revolves around the tool tip (**Figure 8B**). We model this relationship by identifying a center-of-rotation (and corresponding rotation radius) for the tool tip, from which we can sample a valid end-effector position and rotation.

5.5.1 RANSAC Algorithm Parameters

We use the same parameters for k, w, d as in the linear model. We sample $n = 3$ points at each iteration, and use the error threshold $\epsilon = 0.25$. We define functions for model parameterization, error metrics, sampling, and variance in the following sections.

5.5.2 Model Parameter Fitting

We define the optimal model parameters for each iteration of RANSAC as the center-of-rotation (and corresponding rotation radius) of that iteration's samples $\mathbf{S}$:

$$\Theta_{\text{rotation}}(\mathbf{S}) = \langle \theta_\mathbf{c}, \theta_r \rangle \tag{10}$$

where $\theta_\mathbf{c}$ is the position of the center-of-rotation that minimizes its distance from the intersection of lines produced from the position and orientation of each correction sample:

$$\theta_\mathbf{c} = \underset{\mathbf{c}}{\arg\min} \sum_{i=1}^{|\mathbf{S}|} D(\mathbf{c}; \mathbf{a_i}, \mathbf{n_i})^2 \tag{11}$$

where $\mathbf{a_i}$ and $\mathbf{n_i}$ are the position and unit direction vectors, respectively, for sample i in $\mathbf{S}$:

$$\mathbf{a}_i = [x_i, y_i, z_i]^T \qquad \mathbf{n}_i = \left(\mathbf{q}_i \cdot [0, 1, 0, 0]^T\right)\cdot \mathbf{q}' \tag{12}$$

Here, $\mathbf{q_1} \cdot \mathbf{q_2}$ refers to the Hamilton product between two quaternions, and $\mathbf{q}'$ is the inverse of the quaternion $\mathbf{q}$:

$$\mathbf{q}' = [w, x, y, z]'^T = [w, -x, -y, -z]^T \tag{13}$$

We solve for the center-of-rotation by adapting a method for identifying the least-squares intersection of lines Traa (2013). We consider each sample i to be a ray originating at the point $\mathbf{a_i}$ and pointing in the direction of $\mathbf{n_i}$. The center-of-rotation of a set of these rays is thus the point that minimizes the distance between itself and each ray. We define this distance as the piecewise function:

$$D(\mathbf{c}; \mathbf{a}, \mathbf{n}) = \begin{cases} \|(\mathbf{c} - \mathbf{a}) - d\cdot\mathbf{n}\|_2 & \text{if } d > 0 \\ \|\mathbf{c} - \mathbf{a}\|_2 & \text{otherwise} \end{cases} \tag{14}$$

where d is the distance between $\mathbf{a}$ and the projection of the candidate centerpoint $\mathbf{c}$ on the ray:

$$d = (\mathbf{c} - \mathbf{a})^T\mathbf{n} \tag{15}$$

We solve for $\theta_\mathbf{c}$ using the SciPy implementation of the Levenberg-Marquardt method for non-linear least-squares optimization, supplying **Equation 14** as the cost function. We then solve for the radius corresponding to $\theta_\mathbf{c}$:

$$\theta_r = \frac{1}{|\mathbf{S}|}\sum_{i=0}^{|S|}\|\mathbf{a_i} - \theta_\mathbf{c}\| \tag{16}$$

5.5.3 Error Function

We define the error of a single data point $\mathbf{p}$ as its distance from the current iteration's center-of-rotation estimate:

$$\delta_{\text{rotation}}(\mathbf{p}, \theta) = \left(\frac{D(\mathbf{c}; \mathbf{a_p}, \mathbf{n_p})}{d_p}\right)^2 \tag{17}$$

Where d_p is defined in **Equation 15**.

5.5.4 Sampling Function

After RANSAC returns the optimal model parameters and corresponding set of inlier points $\hat{\mathbf{I}} \subset \mathbf{C}$, the rotation component of the transformation is first sampled using the "average" rotation $\bar{\mathbf{q}}_c$ from $\hat{\theta}_\mathbf{c}$ to all inlier points:

$$\bar{\mathbf{q}}_c = \arg\max_{\mathbf{q}\in\mathbb{S}^3} \mathbf{q}^T M \mathbf{q} \qquad M = \frac{1}{|\hat{\mathbf{I}}|}\sum_{\mathbf{p}\in\hat{\mathbf{I}}} \mathbf{r_p}\mathbf{r_p}^T \tag{18}$$

Where $\mathbf{r_p}$ is the quaternion rotation between $\hat{\theta}_\mathbf{c}$ and the position of $\mathbf{p}$, defined by normalizing the quaternion consisting of the scalar and vector parts:

$$\mathbf{r_p} = \langle \|\mathbf{a}\|^2 + \mathbf{b}\mathbf{a}^T, \mathbf{b}^T \times \mathbf{a}\rangle \tag{19}$$

$$\mathbf{a} = \mathbf{p}_t - \hat{\theta}_\mathbf{c} \qquad \mathbf{b} = [\|\mathbf{a}\|, 0, 0] \tag{20}$$

The optimal $\bar{\mathbf{q}}_\mathbf{c}$ is the eigenvector corresponding to the largest eigenvalue of M; this represents the sampled rotation from $\hat{\theta}_\mathbf{c}$.

We then sample $\bar{\mathbf{t}}$ by projecting the point at distance $\hat{\theta}_r$ from $\hat{\theta}_\mathbf{c}$ in the direction of $\bar{\mathbf{q}}_\mathbf{c}$:

$$\bar{\mathbf{t}} = \hat{\theta}_\mathbf{c} + \left[\left(\bar{\mathbf{q}}_\mathbf{c} \cdot \left[0, \hat{\theta}_r, 0, 0\right]^T\right) \cdot \bar{\mathbf{q}}_\mathbf{c}{}'\right]_{1..3} \tag{21}$$

Where $\mathbf{x}_{1..3}$ indicates the 3×1 vector obtained by ommitting the first element of a 4×1 vector $\mathbf{x}$. Finally, we return the sample consisting of the translation $\bar{\mathbf{t}}$ and the normalized rotation $\bar{\mathbf{q}}$ between $\bar{\mathbf{t}}$ and $\hat{\theta}_\mathbf{c}$:

$$\Psi(\hat{\mathbf{I}}, \hat{\theta})_{\text{rotation}} = \left\langle \frac{\bar{\mathbf{q}}}{\|\bar{\mathbf{q}}\|}, \bar{\mathbf{t}} \right\rangle \quad \bar{\mathbf{q}} = \langle \hat{\theta}_r\|\mathbf{a}\| + \mathbf{b}\mathbf{a}^T, \mathbf{b}^T \times \mathbf{a}\rangle \quad \mathbf{a} = \hat{\theta}_\mathbf{c} - \bar{\mathbf{t}} \quad \mathbf{b} = \left[\hat{\theta}_r, 0, 0\right] \tag{22}$$

5.6 Best-Fit Model Selection

The linear and rotational tool transform models represent two different relationships between the translational and rotational components of corrections. We now define a metric for selecting between these two models based on how well they fit the correction data:

$$\Psi(\mathbf{C})_{\text{best-fit}} = \begin{cases} \Psi\left(\hat{\mathbf{I}}_l, \hat{\theta}_l\right)_{linear} & \text{if } \Delta_{linear} < \Delta_{rotation} \\ \Psi\left(\hat{\mathbf{I}}_r, \hat{\theta}_r\right)_{\text{rotation}} & \text{otherwise} \end{cases} \tag{23}$$

Where $\hat{\mathbf{I}}_l, \hat{\theta}_l, \hat{\mathbf{I}}_r, \hat{\theta}_r$ represent the optimal inlier points and parameter values from the linear and rotational models, respectively. The fit of the linear model is calculated as its range of values $\mathbf{z}$ projected in the model's 1D space:

$$\Delta_{linear} = \text{range}(\mathbf{z}) \qquad \mathbf{z} = \left\{\left(\mathbf{p}_t - \hat{\theta}_\mu\right)^T \hat{\theta}_\mathbf{u} | \forall \mathbf{p} \in \hat{\mathbf{I}}\right\} \tag{24}$$

The fit of the rotational model is calculated as the range of unit vectors in the direction of each inlier point as measured from the center-of-rotation:

$$\Delta_{rotation} = 1 - \frac{1}{|\hat{\mathbf{I}}|}\left\|\sum_{p\in\hat{\mathbf{I}}}\left[\left(\mathbf{r_p} \cdot [0,1,0,0]^T\right) \cdot \mathbf{r_p}'\right]_{1..3}\right\|_2 \tag{25}$$

where $\mathbf{r_p}$ is defined in **Equation 19**.

5.7 Evaluation

We evaluated the transfer by correction algorithm results on a 7-DOF Jaco2 arm equipped with a two-fingered Robotiq 85 gripper and mounted vertically on a table-top surface (**Figure 9D**). Each evaluation configuration consisted of one task that was 1) demonstrated using the original, *"source"* tool, and 2) corrected to accommodate a novel, *replacement* tool. We describe data collection for each of these steps in the following sections.

5.8 Demonstrations

Three tasks (**Figure 9**) were demonstrated using three prototypical, "source" tools (**Figures 10A–C**), resulting in a total of nine demonstrations. Demonstrations began with the arm positioned in an initial configuration, and with the gripper already grasping the tool. Each tool's grasp remained consistent across all three tasks. Objects on the robot's workspace were reset to the same initial position before every demonstration. We provided demonstrations by indicating keyframes (Akgun et al., 2012) along the trajectory, each of which was reached by moving the robot's arm to the intermediate pose. At each keyframe, the 7D end effector pose was recorded; note that this is the pose of the joint holding the tool, and *not* the pose of the tool-tip itself (since the tool-tip is unknown to the robot). We provided one keyframe demonstration for each combination of tasks and source tools in this manner, each demonstration consisting of 7–12 keyframes (depending on the source tool used) for the sweeping task, 10–11 keyframes (depending on the source tool used) for the hooking task, and 7 keyframes for the hammering task.

We represented each demonstration using a Dynamic Movement Primitive (DMP) (Schaal, 2006; Pastor et al., 2009). A DMP is trained over a demonstration by perturbing a linear spring-damper system according to the velocity and acceleration of the robot's end-effector at each time step. By integrating over the DMP, a trajectory can then be generated that begins at the end-effector's initial position and ends at a specified end point location. Thus, after training a DMP, the only parameter required to execute the skill is the desired end point location. By parameterizing the end point location of each DMP skill model according to object locations, the overall task can be generalized to accommodate new object configurations. We re-recorded the demonstration if the trained DMP failed to repeat the demonstration task with the source tool.

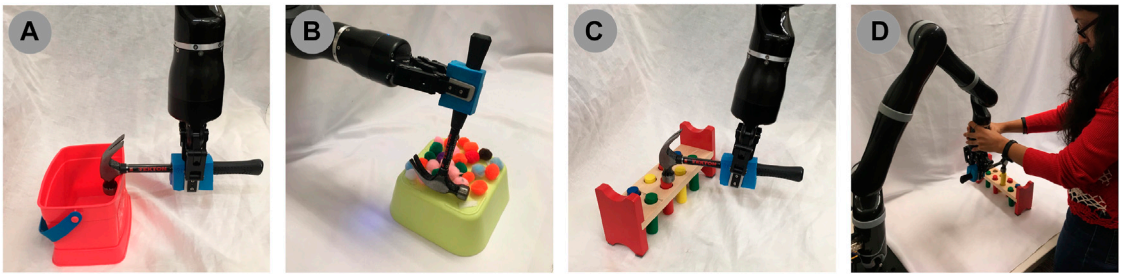


FIGURE 9 | (A) hooking task, **(B)** sweeping task, **(C)** hammering task, and **(D)** the experimental setting.

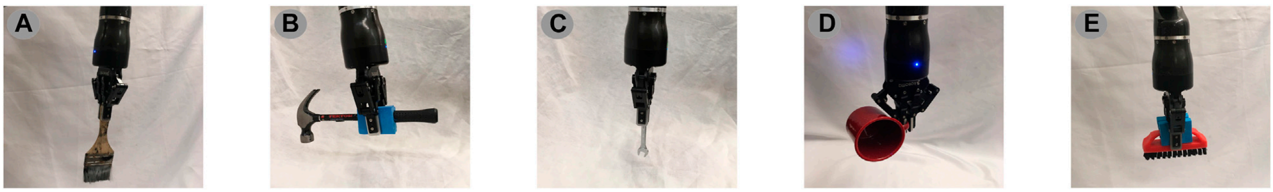


FIGURE 10 | Tools **(A–C)** were used to demonstrate the three tasks shown in **Figure 9**, later transferred to use tools **(D,E)**. These tools exhibit a wide range of grasps, orientations, dimensions, and tooltip surfaces.

5.9 Corrections

Following training, the arm was reset to its initial configuration, with the gripper already grasping a new tool (**Figures 10D,E**). Note that these replacement objects have several surfaces that could be utilized as a tooltip (depending on the task). For example, any point along the rim of the mug (**Figure 10D**) would serve as the prototypical tooltip during a scooping or pouring task. In the context of the hooking and hammering tasks used in our evaluation, however, the bottom of the mug serves as a tooltip. Alternatively, the side of the mug provides a broad surface to perform the sweeping task. This range of potential tooltips on a single object highlights the benefit of using corrections to learn task-specific tooltips, rather than assume that a prototypical tooltip is appropriate for all tasks.

Objects on the robot's workspace were reset to the same initial position as in the demonstrations; this allowed us to ensure that any corrections were made as a result of the change in tool, rather than changes in object positions. The learned model was then used to plan a trajectory in task-space, which was then converted into a joint-space trajectory using TracIK (Beeson and Ames, 2015) and executed, pausing at intervals defined by the keyframe timing used in the original demonstration. When execution was paused, it remained paused until the arm pose was confirmed. If no correction was necessary, the pose was confirmed immediately; otherwise, the arm pose was first corrected by moving the arm to the correct position. Note that this form of corrections assumes that each keyframe constitutes a statically stable state. For tasks involving unstable states, another form of interaction may be used to provide post-hoc corrections, such as critiques (Cui and Niekum, 2018).

Two poses were recorded for each correction: 1) the original end-effector pose the arm attempted to reach (regardless of whether the goal pose was reachable with the new tool), and 2) the end-effector pose following confirmation (regardless of whether a correction was given). Trajectory execution then resumed from the arm's current pose, following the original task-space trajectory so that pose corrections were not propagated to the rest of the trajectory. This process continued until all keyframes were corrected and executed, resulting in the correction matrix **C** (**Equation 2**).

5.10 Measures

For each transfer execution, we measured performance according to a metric specific to the task:

- *Sweeping:* The number of pom-poms swept off the surface of the yellow box.
- *Hooking:* The final distance between the box's target position and the closest edge of the box (measured in centimeters).
- *Hammering:* A binary metric of whether the peg was pressed any lower from its initial position.

5.11 Results

We highlight two categories of results: *Within*-task and *across*-task performance.

5.11.1 Within-Task Transfer

Within-task performance measures the algorithm's ability to model the corrections and perform the corrected task successfully. Transfer was performed using the transform model learned from corrections

Performance Threshold	95%	85%	75%	65%	55%	45%
Transfer Executions	61%	72%	72%	72%	83%	89%
Untransformed Executions	11%	11%	11%	11%	11%	11%

FIGURE 11 | Percentage of within-task transfer executions (selected by best-fit model) and untransformed trajectories achieving various performance thresholds (defined as the % of maximum performance metric for that task, described in **Section 5.10**). Our proposed models result in a higher percentage of transfer executions that complete the task to a high performance threshold (e.g., sweeping ≥ 85% of the objects off the table). Furthermore, while the untransformed baseline produces all-or-nothing performance behavior, our models degrade gracefully, resulting in partial task completion (represented by lower % performance thresholds) even when the learned transform is non-optimal.

on *that same tool-task pairing*. For example, for the sweeping task model learned using the hammer, corrections were provided on the replacement tool (e.g., a mug) and then used to perform the sweeping task using that same mug. For each source tool, we evaluated performance on all three tasks using each of the two replacement objects, resulting in 18 sets of corrections (one for each combination of task, source tool, and replacement tool) per tool transform model (linear and rotational).

Using the better-performing model resulted in ≥ 85% of maximum task performance in 83% of cases. The better-performing model was selected using the best-fit metric in 72% of cases. **Figure 11** lists the percentage of transfer executions (using the best-fit model) that achieve multiple performance thresholds, where best-fit results were recorded as the performance of the model returned by **Equation 23**.

We scaled the result of each transfer execution between 0 and 1, with 0 representing the initial state of the task and 1 representing maximum performance according to the metrics in **Section 5.10**. **Figure 12** reports the performance distribution aggregated over all tasks, transferred from each of the three source tools to either the scrub-brush (**Figure 10E**, results in **Figure 12A**) or mug (pictured in **Figure 10D**, results in **Figure 12B**) as the replacement tool. The mean performance results are reported in **Figure 13A**, with darker cells indicating better performance. Overall, the transform returned using the best-fit metric resulted in average performance of 6.9x and 5.9x that of the untransformed trajectory when using the scrub-brush and mug, respectively, as replacement tools.

5.11.2 Across-Task Transfer

Across-task transfer performance measures the generalizability of corrections learned on one task when applied to a *different* task using the same tool, without having received any corrections on that tool-task pairing. For example, the hooking task was learned using the hammer, and transferred to the mug using corrections obtained on the sweeping task. We evaluated 36 total transfer executions (one per combination of demonstration task, source tool, correction task (distinct from the demonstration task), and replacement tool) per tool transform model (linear and rotational).

Figure 14 reports the performance distribution aggregated over all tasks, transferred from each of the three source tools to either the scrub-brush (**Figure 14A**) or mug (**Figure 13B**) as the replacement tool. The mean performance results are reported in **Figure 13B**, with darker cells indicating better performance. Overall, the transform returned using the best-fit metric resulted in average performance of 1.6*x* and 0.94*x* that of the untransformed trajectory when using the scrub-brush and mug, respectively, as replacement tools. The performance distribution is improved when using the transform learned from corrections, resulting in 2.25*x* as many task executions achieving ≥ 25% of maximum task performance.

In order to understand the conditions under which a transform can be reused successfully in the context of another task, we also report the mean performance results for a subset of the across-task executions (**Figure 13C**). This subset consists of only the task executions where the relative orientation is the same between 1) the source tool's tooltips used for the source and target

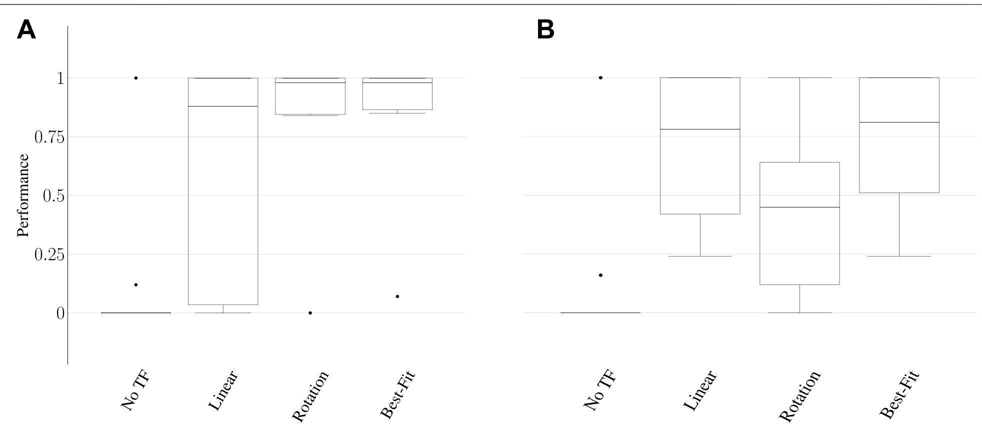


FIGURE 12 | Aggregate performance results for within-task transfer using the scrub-brush **(A)** and mug **(B)** as the replacement tool. Performance was measured for each task according to the metrics in **Section 5.10**, and are scaled between 0–1. These results highlight the need for multiple tool transform models; while both models greatly outperform the baseline task performance (when no transform is used), note that neither model results in the *best* performance over all tasks and replacement tools. Using the best-fit metric to select the more appropriate model for each tool-task pairing resulted in the best overall performance.

A

Replacement Tool	No TF	Linear	Rotation	Best-fit
Brush	0.124	0.648	0.850	**0.863**
Mug	0.129	0.738	0.489	**0.765**

B

Replacement Tool	No TF	Linear	Rotation	Best-fit
Brush	0.124	0.085	0.184	**0.203**
Mug	**0.129**	0.128	0.080	0.121

C

Replacement Tool	No TF	Linear	Rotation	Best-fit
Brush	0.024	0.053	0.268	**0.302**
Mug	0.097	0.162	0.101	**0.162**

FIGURE 13 | Mean performance of **(A)** within-task and **(B)** across-task transfer to the brush and mug replacement tools over all 18 transfer executions for each tool. **(C)** Mean performance of across-task transfer to the brush and mug replacement tools over the subset of transfer executions in which the transformation between source and correction tasks is similar for the source and replacement tool (10 executions for the brush, 12 for the mug). Darker cells indicate higher average performance.

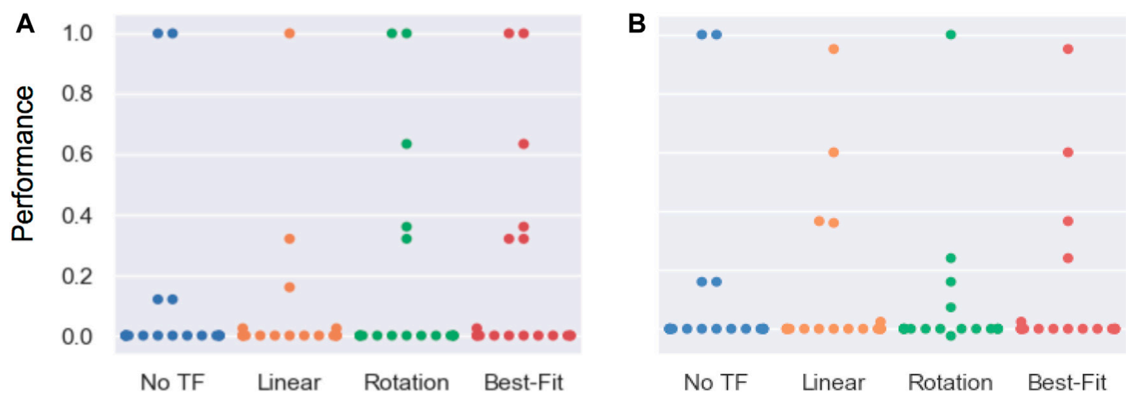


FIGURE 14 | Results for across-task transfer using the scrub-brush **(A)** and mug **(B)** as the replacement tool. Performance was measured according to the metrics in **Section 5.10**, scaled between 0–1. These results represent the generalizability of a transform model learned on one task and then applied to a different task using the same tool. Each point represents the performance of a single transfer execution.

tasks and 2) the replacement tool's tooltips used for the same two tasks. This subset consisted of 10 executions for the scrub-brush, and 12 for the mug. Overall, for this subset of executions, the transform returned using the best-fit metric resulted in average performance of 12.6x and 1.7x that of the untransformed trajectory when using the scrub-brush and mug, respectively, as replacement tools.

5.12 Discussion

Our within-task transfer evaluation tested whether we can model the transform between two tools in the context of the same task (represented by the solid blue arrow in **Figure 15**) using corrections. Our results indicate that one round of corrections typically is sufficient to indicate this relationship between tools; collectively, the linear and rotational models achieved ≥ 85% of maximum task performance in 83% of cases. Individually, the models selected by the best-fit metric achieved this performance threshold in 72% of cases. This indicates that, in general, the fit of the model itself can be used to indicate the relationship between end-effector position and orientation for a given tool/task combination.

Aside from analyzing high task performance, we are also interested in whether our approach enables graceful degradation; even if the robot is unable to complete the task fully with a new tool, ideally it will still have learned a transform that enables *partial* completion of the task. The results shown in **Figure 11** demonstrate that Transfer by Correction offers robust behavior such that even when it results in sub-optimal performance, it still meets lower performance thresholds in nearly 90% of cases. In contrast, the untransformed baseline does not meet lower performance thresholds, and thus produces all-or-nothing results that lack robustness.

The primary benefit of modeling corrections (as opposed to re-learning the task for the new tool) is two-fold: First, the robot learns a transformation that reflects *how* the task has changed in response to the new tool, which is potentially generalizable to other tasks (as we discuss next). We hypothesize that in future work, this learned transform could be parameterized by features of the tool (after corrections on multiple tools). Second, since we do not change the underlying task model, but instead apply the learned transform to the resulting trajectory, the underlying task model is left unchanged. We expect that this efficiency benefit

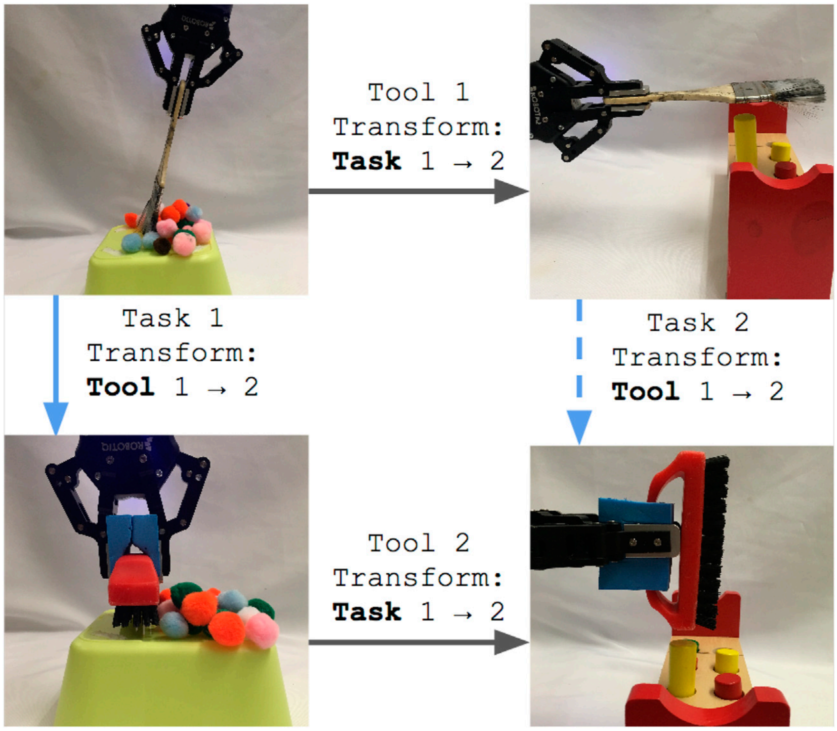


FIGURE 15 | Corrections indicate the transform from tool 1 to tool 2 for the same task (indicated by the solid blue arrow). Our within-task transfer evaluation tested whether we can use corrections to sufficiently model this relationship. Different tasks may use different tooltips from the same tool (such as the different tooltips used to complete tasks 1 and 2). Our across-task evaluation tests whether the transform learned from corrections (solid blue arrow) can be reused as the transform between the two tools for another task (indicated by the dashed blue arrow).

would be most evident when transferring a more complex task model trained over many demonstrations; rather than require more demonstrations with the new tool in order to re-train the task model, the transform would be applied to the result of the already-trained model.

We have also explored how well this transform generalizes to other tasks. Different tooltips on the same tool may be used to achieve different tasks, such as how the end and base of the paintbrush are used to perform sweeping and hammering tasks, respectively, in **Figure 15**. While we do not explicitly model the relationship between tooltips on the same tool (represented by the top grey arrow in **Figure 15**), they are inherent to the learned task models. A similar relationship exists for the replacement tool (represented by the bottom grey arrow in **Figure 15**). Our across-task evaluation seeks to answer whether the relationship between tools in the context of the first task (solid blue arrow) can be reused for a second task (represented by the dashed blue arrow) without having received any corrections on that tool/task combination (tool 2 and task 2). While we see lower performance in across-task evaluations compared to the within-task evaluations, it does improve transfer in 27.8% of across-task transfer executions (in comparison to the untransformed trajectory).

In the general case, our results also indicate that we cannot necessarily reuse the learned transformation on additional tasks, as average performance in across-task transfer is slightly worse than that of the untransformed trajectory when the mug is used as a replacement tool. This presents the question: Given a transform between two tools in the context of one task, under what conditions can that transform be reused in the context of another task *without additional corrections or training*? We do see that across-task performance is best when considering only the subset of cases where the relationship between the tooltips used in either task is similar for the source and replacement tools (in our evaluation, this is 10 of 18 executions using the brush, and 12 of 18 executions using the mug). Within this subset, across-task transfer improves performance in 41% of transfer executions. From this we draw two conclusions: 1) the transform applied to a tool is contextually dependent on the source task, target task, and tooltips of the source and replacement tool, and 2) a transform can be reused when the relationship between tooltips used in either task is similar for the source and replacement tools.

Overall, our evaluation resulted in the following key findings:

Insight #1: Corrections provide a sample of the *constrained* transform between the tooltip and the robot's end-effector. This underlying constraint is task-dependent; our best-fit model results indicate that **multiple constraint types should be modeled and evaluated for each task**, with the best-fitting model used to produce the final transform output.

Insight #2: While the tooltip transform is task-specific, it can be applied to additional tasks under certain conditions. This is dependent on a second transform: the transform between *multiple tooltips* on the same tool. **A tooltip transform can be reused for an additional task when the transform between the tooltips used to complete 1) the corrected task and 2) the additional task are similar for the two tools.**

6 CONCLUSION

Tool use is a hallmark of human cognition and tool improvisation is a characteristic of human creativity. As robots enter human society, we expect human-like tool improvisation from robots as well. This paper makes three contributions to robot creativity in using novel tools to accomplish everyday tasks. First, it presents a high-level decomposition of the task of tool improvisation into a process of tool exploration, tool evaluation, and adaptation of task models to the novel tool. Second, it demonstrates the importance of *tooltip constraints* in guiding successful tool use throughout this process. Third, it describes a method of learning by correction: repeating a known task with an unknown tool in order to record a human teacher's corrections of the robot's motion.

We focused on how the relationship between the robot's gripper and the tooltip dictates how the robot's action model should be adapted to the new tool. A challenge in identifying this relationship is that 1) there are many candidate tooltips on each tool, and 2) for each tooltip, there exists a one-to-many relationship between the tooltip and end-effector poses that fulfill the tooltip constraint.

In this paper, we validated this one-to-many mapping through a simulated experiment in which we demonstrate a relationship between pose variations and task performance. Our experimental results indicate that the sensitivity of tooltip constraints depends on the surface of the tool being used, and that as the tool pose deviates from these constraints, the resulting effect on task performance is nonlinear.

We then examined the opposite mapping: A many-to-one mapping between pose feedback provided by a human teacher, and the optimal, underlying tooltip constraint. We developed the Learning by Correction algorithm, and demonstrated that a human teacher can indicate the tooltip constraints for a specific tool-task pairing by correcting the robot's motion when using the new tool. We modeled the underlying tooltip constraint in two ways, using a linear and rotation model, and also present a metric for choosing the better-fitting model for a set of corrections. We demonstrated how this model of the tooltip constraint can then be used to successfully plan and execute the task using that tool with high task performance in 83% of task executions. We also explored how this tooltip constraint model can be generalized to additional tasks using the same novel tool, without requiring any additional training data.

Overall, we expect that a focus on identifying novel tools, evaluating novel tools, and adapting task models to novel tools in accordance to tooltip constraints is essential for enabling creative tool use. Our results indicate that successful task adaptation for a new tool is dependent on the tool's usage within that task, and that the transform model learned from interactive corrections can be generalized to other tasks providing a similar context for the new tool. Put together, these results provide a process account of robot creativity in tool use (tool identification, evaluation and adaptation), a content account (highlighting the importance of tooltips), as well as an algorithmic account of learning by correction.

6.1 Open Questions

In this paper, we have presented a corrections-based approach to sampling and modeling the transform resulting from a tool replacement. In doing so, we model a single, *static* transform for a particular tool/task pairing. We have evaluated how well this model transfers to other tasks using the same tool replacement. An extension of this work would consider transfer across *tools*.

We envision that a robot could not only model the transform samples obtained by interactive corrections, but also learn to generalize that model to other, similar tools. For example, after receiving corrections for one ladle for a scooping task, the robot would ideally be able to model those corrections such that it would apply to ladles of different shapes or proportions as well. We anticipate that a robot could learn an underlying relationship between visual object features (such as dimensions or concavity) and the resulting transform for that tool.

Meta-learning has been successfully applied to learning problems in computer vision domains and fully-simulated reinforcement learning problems (Duan et al., 2017; Chelsea et al., 2017). When applied to the domain of tool transfer, meta-learning would ideally enable a robot to use extensive background training to learn the common relationships between visual features and tooltips that are shared by tools within their respective categories (e.g., cups, knives, scoops). When presented with a novel category of tools, the robot would then only need demonstrations using a small number of tools within the new category in order to learn the relationship between visual features and tooltips within that category. However, as demonstrated in this paper, tooltips are task-specific; within a single tool, the tooltip used to complete one task (e.g., the surface of a hammer used to hammer a nail) is not necessarily the same as the tooltip used to complete another task (e.g., the side of the hammer may be used to sweep objects off a surface, or the claw-end of the hammer may be used to remove a nail). This lack of task-specific training data presents a challenge for future work, as relying on a dataset containing a single, canonical tooltip for each tool would fail to capture the task-contextual nature of tool use.

Finally, this paper has explored one method of interaction to enable a human teacher to provide corrections to the robot. However, in human-in-the-loop learning problems, the ideal interaction type is dependent on the teacher's role in the learning system, and the context in which the robot is used (Cui et al., 2021). For example, the teacher may not have time to correct every step of the robot's action, or may instead prefer to provide corrections only after the robot has tried and failed to complete a task. We anticipate that future work may enable a robot to obtain correction data from a broader set of interaction types.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

This paper is based on the PhD dissertation of TF, with AG and AT as advisors.

FUNDING

This material is based on work supported in part by Office of Naval Research grants N00014-18-1-2503 and N00014-14-1-0120, and the IBM PhD Fellowship.

ACKNOWLEDGMENTS

Section 5 is based upon the authors' previously published work at the International Conference on Autonomous Agents and Multiagent Systems (Fitzgerald et al., 2019). The authors would like to thank Jinqi Chen for configuring the simulated robot environment and tasks, and for her insights in analyzing the simulated dataset. We would also like to thank Elaine Short for her valuable insights in designing the tool transform models, and Sonia Chernova for many helpful discussions in planning the physical robot evaluation.

REFERENCES

Agostini, A., Aein, M. J., Szedmak, S., Aksoy, E. E., Piater, J., and Wörgötter, F. (2015). "Using Structural Bootstrapping for Object Substitution in Robotic Executions of Human-like Manipulation Tasks," in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, September 2015 (IEEE), 6479–6486. doi:10.1109/iros.2015.7354303

Akgun, B., Cakmak, M., Jiang, K., and Thomaz, A. L. (2012). Keyframe-based Learning from Demonstration. *Int. J. Soc. Robot.* 4, 343–355. doi:10.1007/s12369-012-0160-0

Argall, B. D., Chernova, S., Veloso, M., and Browning, B. (2009). A Survey of Robot Learning from Demonstration. *Robot. Auton. Syst.* 57, 469–483. doi:10.1016/j.robot.2008.10.024

Argall, B. D., Sauser, E. L., and Billard, A. G. (2010). "Tactile Guidance for Policy Refinement and Reuse," in 2010 IEEE 9th International Conference on Development and Learning (ICDL), Ann Arbor, Michigan, August 2010 (IEEE), 7–12. doi:10.1109/devlrn.2010.5578872

Bajcsy, A., Losey, D. P., O'Malley, M. K., and Dragan, A. D. (2018). "Learning from Physical Human Corrections, One Feature at a Time," in Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (ACM), Chicago, Illinois, March 2018, 141–149. doi:10.1145/3171221.3171267

Beeson, P., and Ames, B. (2015). "Trac-Ik: An Open-Source Library for Improved Solving of Generic Inverse Kinematics," in 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), Seoul, South Korea, November 2015 (IEEE), 928–935. doi:10.1109/humanoids.2015.7363472

Bird, J., and Stokes, D. (2006). "Evolving Minimally Creative Robots," in Proceedings of the Third Joint Workshop on Computational Creativity, Riva del Garda, Italy, August 2006 (Amsterdam: IOS Press), 1–5.

Brown, S., and Sammut, C. (2012). "Tool Use and Learning in Robots," in *Encyclopedia of the Sciences of Learning* (Springer), 3327–3330. doi:10.1007/978-1-4419-1428-6_1652

Chelsea, F., Yu, T., Zhang, T., Abbeel, P., and Levine, S. (2017). *One-shot Visual Imitation Learning via Meta-Learning*. Mountain View, California: arXiv. preprint arXiv:1709.04905.

Chernova, S., and Thomaz, A. L. (2014). Robot Learning from Human Teachers. *Synth. Lectures Artif. Intell. Machine Learn.* 8, 1–121. doi:10.2200/s00568ed1v01y201402aim028

Choi, D., Langley, P., and To, S. T. (2018). "Creating and Using Tools in a Hybrid Cognitive Architecture," in AAAI Spring Symposia, Palo Alto, California, March 2018.

Coleman, D., Şucan, I. A., Chitta, S., and Correll, N. (2014). Reducing the Barrier to Entry of Complex Robotic Software: A MoveIt! Case Study. *J. Software Eng. Robotics* 5 (1), 3–16. doi:10.6092/JOSER_2014_05_01_p3

Cui, Y., Koppol, P., Admoni, H., Niekum, S., Simmons, R., Steinfeld, A., et al. (2021). "Understanding the Relationship between Interactions and Outcomes in Human-In-The-Loop Machine Learning," in Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Montreal, QC, Canada. doi:10.24963/ijcai.2021/599

Cui, Y., and Niekum, S. (2018). "Active Reward Learning from Critiques," in 2018 IEEE international conference on robotics and automation (ICRA), Brisbane, QLD, Australia, May 2018 (IEEE), 6907–6914. doi:10.1109/icra.2018.8460854

Dabbeeru, M. M., and Mukerjee, A. (2011). Discovering Implicit Constraints in Design. *Aiedam* 25, 57–75. doi:10.1017/s0890060410000478

Do, T.-T., Nguyen, A., and Reid, I. (2018). "Affordancenet: An End-To-End Deep Learning Approach for Object Affordance Detection," in 2018 IEEE international conference on robotics and automation (ICRA), Brisbane, QLD, Australia, May 2018 (IEEE), 5882–5889. doi:10.1109/icra.2018.8460902

Duan, Y., Andrychowicz, M., Stadie, B., Ho, O. J., Schneider, J., Sutskever, I., et al. (2017). "One-shot Imitation Learning," in *Advances in Neural Information Processing Systems* Long Beach, California: Curran Associates, Inc., 1087–1098.

Dym, C. L., and Brown, D. C. (2012). *Engineering Design: Representation and Reasoning*. Cambridge University Press.

Fang, K., Zhu, Y., Garg, A., Kurenkov, A., Mehta, V., Fei-Fei, L., et al. (2018). "Learning Task-Oriented Grasping for Tool Manipulation from Simulated Self-Supervision," in Proceedings of Robotics: Science and Systems, Pittsburgh, Pennsylvania, June 2018 (Pittsburgh, Pennsylvania). doi:10.15607/RSS.2018.XIV.012

Fauconnier, G., and Turner, M. (2008). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.

Fischler, M. A., and Bolles, R. C. (1981). Random Sample Consensus. *Commun. ACM* 24, 381–395. doi:10.1145/358669.358692

Fitzgerald, T., Goel, A., and Thomaz, A. (2018). Human-guided Object Mapping for Task Transfer. *J. Hum. Robot. Interact.* 7, 1–24. doi:10.1145/3277905

Fitzgerald, T., Goel, A., and Thomaz, A. (2017). "Human-robot Co-creativity: Task Transfer on a Spectrum of Similarity," in International Conference on Computational Creativity (ICCC), Atlanta, Georgia, June 2017.

Fitzgerald, T., Short, E., Goel, A., and Thomaz, A. (2019). "Human-guided Trajectory Adaptation for Tool Transfer," in International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Montreal, Quebec, Canada, May 2019 (International Foundation for Autonomous Agents and Multiagent Systems), 1350–1358.

Fu, J., Levine, S., and Abbeel, P. (2016). "One-shot Learning of Manipulation Skills with Online Dynamics Adaptation and Neural Network Priors," in 2016 IEEE/RSJ International Conference onIntelligent Robots and Systems (IROS), Daejeon, South Korea, October 2016 (IEEE), 4019–4026. doi:10.1109/iros.2016.7759592

Gajewski, P., Ferreira, P., Bartels, G., Wang, C., Guerin, F., Indurkhya, B., et al. (2018). "Adapting Everyday Manipulation Skills to Varied Scenarios," in 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, May 2019. IEEE, 1345–1351.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception* Boston: Houghton Mifflin.

Goel, A. K. (1997). Design, Analogy, and Creativity. *IEEE Expert* 12, 62–70.

Goel, A. K., Fitzgerald, T., and Parashar, P. (2020). "Analogy and Metareasoning: Cognitive Strategies for Robot Learning," in *Human-Machine Shared Contexts* (Elsevier), 23–44. doi:10.1016/b978-0-12-820543-3.00002-x

Gopinath, D., and Weinberg, G. (2016). A Generative Physical Model Approach for Enhancing the Stroke Palette for Robotic Drummers. *Robot. Auton. Syst.* 86, 207–215. doi:10.1016/j.robot.2016.08.020

Gubenko, A., Kirsch, C., Smilek, J. N., Lubart, T., and Houssemand, C. (2021). Educational Robotics and Robot Creativity: An Interdisciplinary Dialogue. *Front. Robot. AI* 8, 178. doi:10.3389/frobt.2021.662030

Hoffmann, H., Chen, Z., Earl, D., Mitchell, D., Salemi, B., and Sinapov, J. (2014). Adaptive Robotic Tool Use under Variable Grasps. *Robot. Auton. Syst.* 62, 833–846. doi:10.1016/j.robot.2014.02.001

Houkes, W., and Vermaas, P. E. (2010). *Technical Functions: On the Use and Design of Artefacts*, Vol. 1. Springer Science & Business Media.

Kemp, C. C., and Edsinger, A. (2006). "Robot Manipulation of Human Tools: Autonomous Detection and Control of Task Relevant Features," in Proceedings of the Fifth International Conference on Development and Learning, Bloomington, Indiana, June 2006.

Kemp, C., Edsinger, A., and Torres-Jara, E. (2007). Challenges for Robot Manipulation in Human Environments [grand Challenges of Robotics]. *IEEE Robot. Automat. Mag.* 14, 20–29. doi:10.1109/mra.2007.339604

Killian, T. W., Daulton, S., Konidaris, G., and Doshi-Velez, F. (2017). "Robust and Efficient Transfer Learning with Hidden Parameter Markov Decision Processes," in *Advances in Neural Information Processing Systems* San Francisco, California: Curran Associates, Inc., 6250–6261.

Kroemer, O., Ugur, E., Oztop, E., and Peters, J. (2012). "A Kernel-Based Approach to Direct Action Perception," in 2012 IEEE international Conference on Robotics and Automation, Saint Paul, Minnesota, May 2012, (IEEE), 2605–2610. doi:10.1109/icra.2012.6224957

Levihn, M., and Stilman, M. (2014). "Using Environment Objects as Tools: Unconventional Door Opening," in 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, Illinois, September 2014 (IEEE), 2502–2508. doi:10.1109/iros.2014.6942903

Markley, F. L., Cheng, Y., Crassidis, J. L., and Oshman, Y. (2007). Averaging Quaternions. *J. Guidance Control Dyn.* 30, 1193–1197. doi:10.2514/1.28949

Myers, A., Teo, C. L., Fermüller, C., and Aloimonos, Y. (2015). "Affordance Detection of Tool Parts from Geometric Features," in IEEE International Conference on Robotics and Automation (ICRA), Seattle, Washington, May 2015 (IEEE), 1374–1381. doi:10.1109/icra.2015.7139369

Nair, L., and Chernova, S. (2020). Feature Guided Search for Creative Problem Solving through Tool Construction. *Front. Robot. AI* 7, 205. doi:10.3389/frobt.2020.592382

Nair, L., Srikanth, N. S., Erickson, Z. M., and Chernova, S. (2019). "Autonomous Tool Construction Using Part Shape and Attachment Prediction," in Proceedings of Robotics: Science and Systems (RSS '19), Freiburg, Germany, June 2019. doi:10.15607/rss.2019.xv.009

Pastor, P., Hoffmann, H., Asfour, T., and Schaal, S. (2009). "Learning and Generalization of Motor Skills by Learning from Demonstration," in IEEE International Conference on Robotics and Automation, 2009. ICRA'09, Kobe, Japan, May 2009 (IEEE), 763–768. doi:10.1109/robot.2009.5152385

Penn, D. C., Holyoak, K. J., and Povinelli, D. J. (2008). Darwin's Mistake: Explaining the Discontinuity between Human and Nonhuman Minds. *Behav. Brain Sci.* 31, 109–130. doi:10.1017/s0140525x08003543

Rozo, L., Jiménez, P., and Torras, C. (2013). A Robot Learning from Demonstration Framework to Perform Force-Based Manipulation Tasks. *Intel Serv. Robot.* 6, 33–51. doi:10.1007/s11370-012-0128-9

Sarathy, V., and Scheutz, M. (2018). Macgyver Problems: Ai Challenges for Testing Resourcefulness and Creativity. *Adv. Cogn. Syst.* 6, 31–44.

Sauser, E. L., Argall, B. D., Metta, G., and Billard, A. G. (2012). Iterative Learning of Grasp Adaptation through Human Corrections. *Robot. Auton. Syst.* 60, 55–71. doi:10.1016/j.robot.2011.08.012

Schaal, S. (2006). "Dynamic Movement Primitives-A Framework for Motor Control in Humans and Humanoid Robotics," in *Adaptive Motion of Animals and Machines* (Springer), 261–280.

Schubert, A., and Mombaur, K. (2013). "The Role of Motion Dynamics in Abstract Painting," in Proceedings of the Fourth International Conference on Computational Creativity (Citeseer), Sydney, Australia, June 2013.

Sinapov, J., and Stoytchev, A. (2008). "Detecting the Functional Similarities between Tools Using a Hierarchical Representation of Outcomes," in 7th IEEE International Conference on Development and Learning, ICDL 2008, Monterey, California, August 2008 (IEEE), 91–96. doi:10.1109/devlrn.2008.4640811

Srinivas, A., Jabri, A., Abbeel, P., Levine, S., and Finn, C. (2018). "Universal Planning Networks," in Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden. arXiv. preprint arXiv:1804.00645.

Taylor, M. E., and Stone, P. (2009). Transfer Learning for Reinforcement Learning Domains: A Survey. *J. Machine Learn. Res.* 10, 1633–1685. doi:10.5555/1577069.1755839

Traa, J. (2013). *Least-squares Intersection of Lines*. Illinois: UIUC.

Vaesen, K. (2012). The Cognitive Bases of Human Tool Use. *Behav. Brain Sci.* 35, 203. doi:10.1017/s0140525x11001452

Vigorito, C. M., and Barto, A. G. (2008). "Hierarchical Representations of Behavior for Efficient Creative Search," in AAAI Spring Symposium: Creative Intelligent Systems, Palo Alto, California, March 2008, 135–141.

Yannakakis, G. N., Liapis, A., and Alexopoulos, C. (2014). *Mixed-initiative Co-creativity* Fort Lauderdale, Florida: International Conference on the Foundations of Digital Games.

Conflict of Interest: The authors declare that this study received funding from IBM in the form of a PhD Fellowship. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 24 November 2021
doi: 10.3389/frobt.2021.631752

Creative AI and Musicking Robots

Craig Vear *

Institute of Creative Technologies, De Montfort University, Leicester, United Kingdom

This article discusses the creative and technical approaches in a performative robot project called "*Embodied Musicking Robots*" (2018–present). The core approach of this project is human-centered AI (HC-AI) which focuses on the design, development, and deployment of intelligent systems that cooperate with humans in real time in a "deep and meaningful way."[1] This project applies this goal as a central philosophy from which the concepts of creative AI and experiential learning are developed. At the center of this discussion is the articulation of a shift in thinking of what constitutes creative AI and new HC-AI forms of computational learning from inside the flow of the shared experience between robots and humans. The central case study (*EMRv1*) investigates the technical solutions and artistic potential of AI-driven robots co-creating with an improvising human musician (the author) in real time. This project is ongoing, currently at v4, with limited conclusions; other than this, the approach can be felt to be cooperative but requires further investigation.

Keywords: music, robots, creativity, human-centered AI, creative AI

OPEN ACCESS

Edited by:
Patricia Alves-Oliveira, University of Washington, United States

Reviewed by:
Caterina Moruzzi, University of Konstanz, Germany
Shiqing He, Texas A&M University, United States

****Correspondence:***
Craig Vear
cvear@dmu.ac.uk

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *20 November 2020*
Accepted: *14 October 2021*
Published: *24 November 2021*

Citation:
Vear C (2021) Creative AI and Musicking Robots. Front. Robot. AI 8:631752. doi: 10.3389/frobt.2021.631752

INTRODUCTION

The Goal

The aim of this practice-based research project was to investigate the technical solutions and artistic potential of AI-driven robots co-creating with a human musician in real time. This research extends and enhances existing research in this area, specifically that of computational creativity (e.g., McCormack and d'Inverno (McCormack and d'Inverno, 2012)), AI and music (e.g., Miranda (Miranda, 2021)), and robotic musicianship (e.g., Bretan et al. (Weinberg et al., 2020)), with a specific focus on the embodied relationship among agent- robot, sound presence, human musician, and the flow of co-creativity, with the aim of enhancing the creativity of humans. This is a rich and emerging area with many solutions which are currently being developed, most of which are dealing with in-the-loop solutions for human–robot music interaction. For example, the cooperative AI at the heart of "*In A Silent Way*" (McCormack et al., 2019) is trained using performance data and communicates with the human musicians through real-time sound generation and emoticons in order to generate a sense of trust. Additionally, in *Design Considerations for Real-Time Collaboration with Creative Artificial Intelligence*, McCormack (McCormack et al., 2020) offers a framework for maximizing the human–AI creative interaction, which can be migrated to human–robotic musicking.

The *Embodied Musicking Robots* (EMR) project contributes to this discussion by asking the following research question:

> If we want robots to join us inside the creative acts of music, then how do we design and develop robot systems that prioritize the relationships that bind musicians inside the flow of music-making?

[1]MIT (2019). Available at: https://hcai.mit.edu (Accessed 2020/10/23).

This question comes from deep and meaningful experiences that I had as a high-level professional musician for over 30 y and support the core goal of HC-AI. As such, its focus is to seek solutions for the stimulation of the relationships generated inside the real-time music-making, which I outlined in detail in Vear (2019) with its basic structure being split into these two domains:

1) *Taking in*: within the flow, musicians make connections with the AI as they reach out, suggest, offer, and shift through the tendrils of affordance experienced through the notions of.
   - *Liveness*: the sensation that the AI is cooperating in the real time making of music, and this meaningful engagement feels "alive."
   - *Presence*: an experience that some*thing* is there or I am *there*.
   - *Interaction*: the interplay of self with environments, agents, and actants.
2) *Taken into*: the AI can establish a world of creative possibilities for exploration through the flow through the domains of.
   - *Play*: the pure play of musicking happens inside a play-sphere in which the idea and musicking are immutably fused.
   - *Time*: the perception of time (of now, past, future, and the meanwhile of multiple convolutions of time) inside musicking plays a central role to the experience of the musician
   - *Sensation*: is an esthetic awareness in the experience of an environment (music world) as felt through their senses.

However, I must stress that the EMR project is ongoing and so far relatively unfunded, with the limitations imposed by COVID and repeated lockdowns, and has only had the author in-the-loop. Therefore, this article should be read as more a hopeful, position statement, with some (autobiographical) evidence to support the authors understanding that this system feels like it is stimulating relationships inside music-making, rather than simulating the movements and sounds of a robotic musician.

Definitions

Before I describe the solutions that I designed and deployed in the *EMR* project; I need to simply define what I mean by the following terms within the context of this project. This defining process also helped consolidate the design and development of the Creative AI and robotic systems with the goals of HC-AI.

Musicking is the creative acts of real-time music-making. Musicking is a term first created by Christopher Small to define a perspective that "to music is to take part" (Small, 1998). Small wrote that "taking part can happen in any capacity" (Small, 1998) such as performing, composing, and listening (and dancing). It crucially means formation through musicking is formed in the relationships that are established within the realm of taking part with agents, sounds, spaces, and presences that are encountered here.

Flow is the experience of musicking from inside the activity. Within the context of this project, the flow of musicking defines how "musicians become absorbed in the music through a sense of incorporation within their environment (the sound world), a shared effort (with the digital, virtual, AI, and robotic agents), and a loss of awareness of their day-to-day wakefulness and bodily self-consciousness (embodiment with their instrument and into their music)" (Vear, 2019).

Embodiment (in music) is the process in musicking of drawing the musician's sound into their bodily sense of being. This presumes that when musicians make music, it is not a process of outputting sound into the world but an embodied experience of becoming the sound they create in the flow of musicking. Equally, it describes the process of the musician reaching out from this sense of becoming and drawing in the sounds of others so they feel their *presence as sound*. This is a dance of sorts: to touch, to feel, to sense, to work with, to play with, and to hide and seek and flirt and subvert with others through the *flow*.

Creativity: I recognize creativity when play turns into invention within the flow of musicking. As a musician, creativity has to be of value and meaning to me. It needs to be "greater than the sum of its parts" (Vear, 2019); (Boden, 2003; Zedan et al., 2008; Iacoboni, 2009; Pearce, 2010; Thomsom and Jaque, 2017; Zedan et al., 2017) and go beyond merely creating music (manufacturing sound using one's skills). It also goes beyond recognizing that something is new, or novel, or that I have innovated in a given situation. Creativity is felt to be fundamentally new—to my mind—and emergent from my playfulness within the flow. It takes effort and needs feeding, and goes beyond "adhering to a list of ingredients and/or instructions within a prescribed situation; emergent creativity—that is, genuinely original—cannot be replicated by simply repeating a set of rules or prescribed circumstances" (Vear, 2019).

Creativity is giving in to a playful situation that *might* return with a creative spark. Creativity is not constant, reliable, or automatic; it needs nurturing with open, generous, and cultivating energy. On the other hand, it can sustain bold and mischievous challenges or seemingly disruptive engagement designed to rail-road ongoing trains of thought, so long as these are still giving in their nature.

In this article, I define three sub-domains of creativity to highlight the human–robot relationships. These are based on my general experience as an improvising musician and are used to identify the types of co-creativity within musicking from the human musician's perception (note: this project does not deal with notions of machine consciousness or perception):

- *Concurrent*: a sense that both agents (human and robot) are playfully inventing in isolation but within the shared flow of musicking
- *Collaborative*: a sense that both agents are contributing to a shared play idea, feeding a sense of collective invention through individual contribution and perspective
- *Co-creative*: a sense that the robot and human agents are collectively inventing through a stimulated sense that each is in inside the other's head. By this, I mean that the robot/AI, as perceived by the human musician, is in the loop with the human, and together, they are inventing on a singular idea, feeding each other's play as if it were one train of thought.

Creative AI not only includes practices that have AI embedded into the process of creation but also encompasses novel AI approaches in the realization and experience of such work. I define AI as the design, development, and deployment of intelligent agents that respond with insights from their environment and perform actions. Each agent is mainly concerned with a rational action within a given *situation* [taken from AI a modern approach]. The focus of behavioral and embodied AI emphasizes the close-coupled relationship between the *situation* that an intelligent agent is operating in and the *behavior* that it exhibits to cope inside such a *situation*. As such, the focus on intelligent behavior is on the coping systems that are required to maintain a balance of existing within such a situated environment.

With these definitions in mind, the goal of the *EMR* project is to design and develop a creative AI system that enhances human musician creativity by stimulating, inspiring, interacting, and cooperating in the flow of embodied live improvised music-making. Therefore, to build a robot driven by a Creative AI system, it must

1) continually improve by learning from humans and
2) create an effective and fulfilling human–robot interaction experience.

THE PROJECT

My hypothesis to the research question posed before involved the design, development, and deployment of a robotic creative AI that would have a presence within the co-creativity of the flow of musicking and not be an AI zombie. This approach reinforces the personal understanding that when a musician enters the world of musicking, the “I” is coping in a very different world of concern than if they were walking down a street. In a sense, “I” becomes a different creature with a different set of priorities and concerns, outlooks, and sensorial inputs than my normal, human wakefulness. The technical and artistic solution for *EMR* focused on a robot that was first and foremost a coping entity in this specific world of concern (the flow of musicking).

The solution was to develop a system based on these three principles, expanded below:

1) Coping: *EMR* needed to cope in real time within the realm of musicking and be present as sound whose movements are embodied within such flow. This required a non-representational approach to how it related to the flow as the coping mechanisms needed to be open and dynamic enough to cooperate in any given musicking realm. Limiting the robot to a single representation of what musicking is, or might be, imposed onto the system by the human designer(s), would only work in a number of instances.
2) Creative AI dataset and experiential learning: these concepts needed to be designed from within the realm of musicking, prioritizing the phenomena of being inside this realm and capturing an essence of what it means to be embodied within the flow. The concept of experiential learning was designed to support this (discussed later).
3) Belief: the robot needs to believe in its view of the musicking world through limitations, embedded esthetics, and behavioral traits, even with glitches and bugs in the system.

From the human-centered artistic perspective, *EMR* needed to address the following:

- The robot was not an extension of the musician but should extend its creativity.
- The robot should not be an obedient dog or responsive insect jumping at my commands or impetus but a playful other.
- It should not operate as a simulation of play but as a stimulation of the human’s creativity.
- It is not a tool to enhance the human’s creativity but a being with presence in the world that they believe to be co-creating with them.
- It should prioritize emergence, surprise, and mischiefbut not expectation.

TECHNICAL SOLUTION

[Not] The Solution

Before I describe my solution, I would like to describe what it is not using relatively well-known examples (NB is not the current state of the art). First, it is not an instrument-performing robot. For example, *TeoTronico* (2012) is a pianist-robot, designed and built by Matteo Suzzi. This robot plays the piano with dynamic control and articulation, moving 53 levers (described as fingers by Suzzi) with “great accuracy and speed” (Prosseda, 2014). In one example on YouTube, it plays a piece composed by Mozart, extremely well. It seems to have sensitivity about its performance, and even though the designers state that it uses MIDI files or be a “mirror pianist” (Prosseda, 2014), it does not sound like it is driven by a standard quantized MIDI file, so some form of human capture was used that stored a human performance as a MIDI file, which *TeoTronico* replayed. Its flow has been prepackaged and then regurgitated. Its sound is in the now, but its musicking is responding neither responding to the now nor to its environment. As such, to achieve the main aim of the *EMR* project, a technical and artistic approach such as this pianist-robot would fail as it simply could not cooperate with the human.

Second, it is not a goal-specific humanoid robot. For example, environmentally aware and goal-cognizant robots such as those being developed for the human–robot World Cup in 2050 (Robocup, 2015) are sophisticated robots employing the AI that make them aware of their world. In general, these systems use computer vision and sensors to navigate through this world; they have real-time awareness of here and now and are interacting with that in their goal to get the ball and score. The problem of using this kind of approach with the *EMR* project is in the nature of an embodied interaction. In musicking, the embodied relationships with other musicians are with the presence of the others as sound and not with them as human flesh. This relationship goes beyond relating to their physical presence and their movement, although it does play a part in varying degrees and at varying times, in my ongoing relationship-building process. So, an *EMR* needs to create relationships with human musicians through its *presence as sound* yet also has some physical presence and movement to inform this. Using these football robots as an analogy, it is not the physical movement of the robot moving toward the ball, or kicking, that creates the relationships required for this project but the relationship with the flow of the movement of the ball. As such, it is not the movement that incites the sound that is being related to in musicking but the presence of that sound in the flow.

The Solution: *EMRv1*

The solution developed for *EMRv1* consists of the following three main concepts.

Coping

The design of *EMR* was informed by two early articles by the robot innovator Rodney Brooks, specifically *Intelligence without Reason* (Brooks, 1991) and *Intelligence without Representation* (Brooks, 1987). In these, he lays out the foundation of his approach to designing and building robots that are first and foremost able to cope and therefore adapt to a dynamically changing environment within the parameters of specific and multiple goals. This research eventually led to his robots being used for space and sea exploration, military and medical application, and the iRobot Roomba vacuum cleaner series. These Roombas are designed with *iAdapt* AI to be "creatures" that cope in a specific world of concern in real time. They neither have a model of representation of their world (such as building a 3D model of the space through computer vision and object analysis) nor do they make one as it goes about its business, but use goals and strategies to cope with whatever that world can throw at it (static furniture, steps, or chairs that get moved 1 day to the next).

Brooks' foundational theories, and observations of my own Roomba, guided the developed for my *EMR* and generated this set of principles (adapted from Brooks (Brooks, 1987)):

- *EMR* must cope in an appropriate musical manner and in a timely fashion, with the dynamic shifts inside the musicking world;
- *EMR* should be robust to the dynamic environment of musicking; it should not fail to minor changes in the properties of the flow of musicking and should behave appropriately to its ongoing perception of the flow;
- *EMR* should maintain multiple goals, changing as required and adapting to its world by capitalizing on creative opportunity;
- *EMR* should do something in the world of musicking; "it should have some purpose in being" (Brooks, 1987).

Creative AI Dataset and Experiential Learning

This project innovated a different approach to computational learning that involved a human-in-the-loop and in-the-groove approach. This experiential learning (EL) approach trained the AI on the job and crucially inside the flow of embodied musicking. Furthermore, the EL process collaborated with a human musician who was equally learning about this new musicking system. This approach supported both the human and the AI to automatically learn and improve from experience.

The EL process (see **Figure 1**) focused on capturing the physical phenomena of an improvising human musician in the flow of creative musicking. The sensing mechanism used 3D depth tracking of the human musician's body using a Kinect sensor (simply x, y, and z movement of both hands, body center, and head) and the fast Fourier Transfer (FFT) analysis of the live sound (fundamental frequency and amplitude).

The resulting dataset reflected the position and rotation of an embodied musicking body in motion together with the amplitude and frequency analysis of the actual sound made by such movement, without preserving the performer's mass, musculature, melodic shape, or music. Thus, the embodied musicking movement is extracted from the performer's body while they are making music; in a poetic sense, the dataset contains the meta-level DNA of musicking without the specifics or a representation of the music or the human. Recorded audio–video capture of the music performance would always anchor the dataset to a specific person and point in time, whereas the meta-level data could become the building blocks for the virtual composition (see **Figure 2**). Data phrases can be edited, treated, and repurposed by the robot's AI again and again without the risk of repetition.

The initial process involved seeding the Creative AI dataset by capturing the live performance of an improvising musician. Once a small set had been generated, this musician then worked with the robot through a series of training sessions, with these new live data being added to the Creative AI dataset. The more they worked together, the more meta-level DNA of their shared creativity would be put back into the dataset system, thereby improving the AI's knowledge base of the shared experience of embodied musicking.

The EL process was used in two ways: first, as a set of raw data that were called upon by the robot AI during a performance (see below), and second, as data for training the four neural networks. These separated the data into four body parts (head, body, right hand, and left hand) and trained a multilayered perceptron neural network using the body parts' x, y, and z data to correlate with its amplitude for each line. Amplitude was decided as being the generator for the neural networks as the proposed application for *EMRv1* was non-idiomatic improvisation, and therefore, sonic impetus was determined to be a more appropriate factor.

The EL approach learns through an embodied interaction inside the flow of musicking. It utilized the meta-level DNA of its improvising partner—the human musician—and extracted elements from the dataset (any data randomly chosen by the system as it is all endowed with meta-level creativity) into its AI processing and then outputs the resultant sound as music. This EL process enhances the dataset through experience by its embodied coping inside the flow of musicking. The human musician perceives meaning in the robot's musicking who in turn cooperates in the making of music (generally perceiving the relationship through one of the perspectives of creative cooperation discussed earlier) and responds with a creative solution through music. This is then captured using the sensing mechanisms and stored back into the Creative AI dataset. Thus, the cycle of EL continues to enhance and improve the dataset and enlarge the creative AI memory bank of deep and meaningful interactions between humans and robots which in turn forms the basis for future interactions.

Belief

It might seem odd to implement belief into the AI of a robot, given that this term usually refers to religious or spiritual faith,

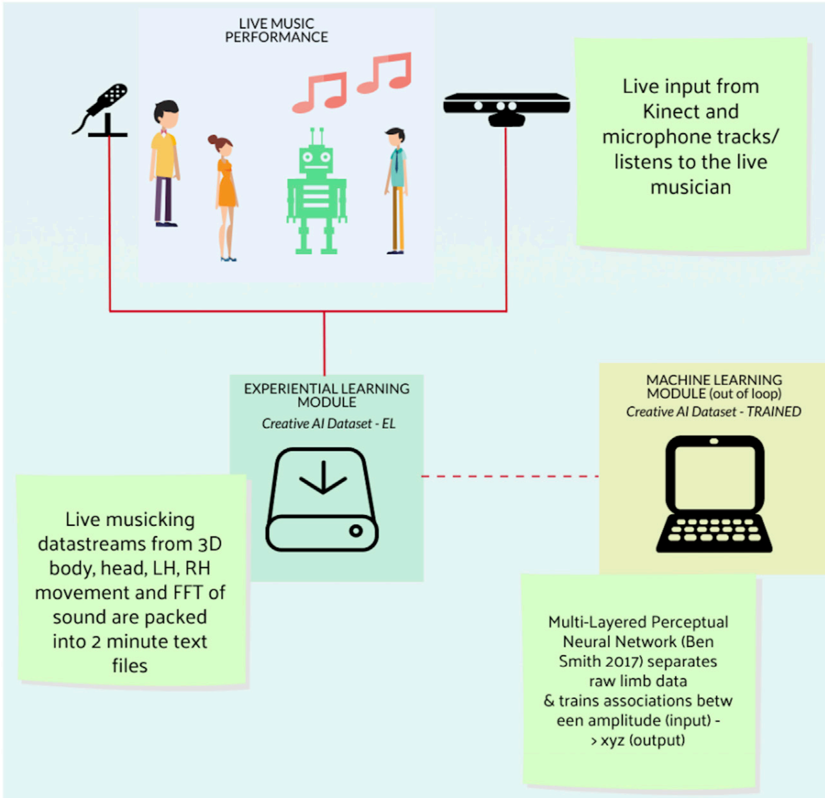


FIGURE 1 | Experiential learning process. This image overviews the basic technical/data structure at play in EMRv1 and also is the foundation for further iterations of EMR.

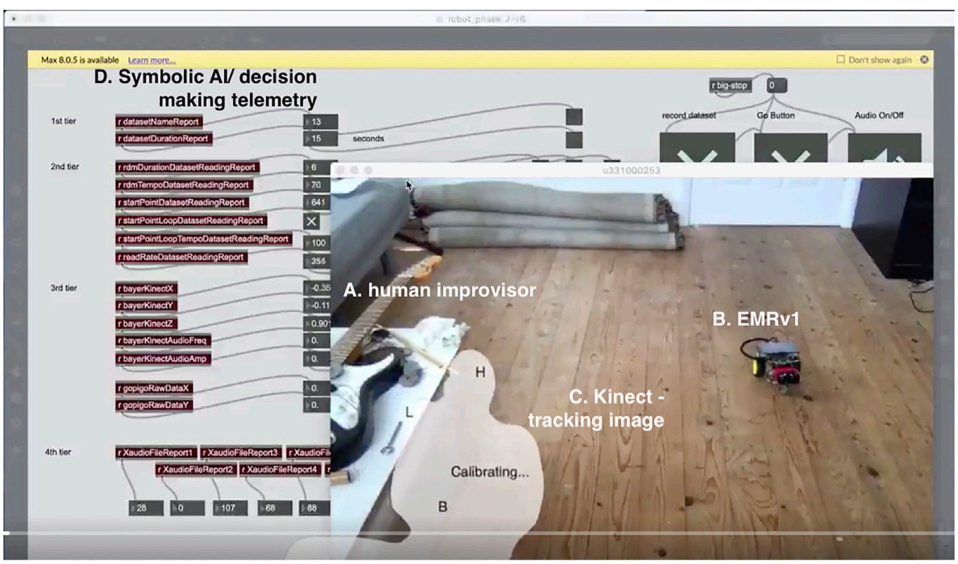


FIGURE 2 | Image of a performance/training session between the author improvising on table-top electric guitar **(A)** and *EMRv1* **(B)**. The image collages the real-time tracking of the Kinect (the ghost image C.), and the data-logging page of Max/MSP patch **(D)**. A video of this session can be found online.[7]

but it is the broader definition of this word that I am particularly interested. Specifically, *EMR* has an acceptance that something is true, or that it has trust or confidence in something from the perspective of the role its belief system plays in the behavior of EMR. I am not suggesting that *EMR* is sentient or has perception of the world but that the robot's operational systems are embedded with structures that it can accept as guiding beliefs. They are:

1) *Movement behavior*: The robot's movement operates within a behavioral system, designed to react openly to the dynamic sound world, and moves the wheels accordingly. The robot AI makes choices determined by whichever goal (listed before) is driving the wheels at any given point but within fixed parameters. The *Embodied Robot for Music* has freedom of choice to operate within such a field of response possibility. These are based on human preferences and outline a range of creative choices which have been determined over several decades. These are personal and subjective, and if these parameters were to be shifted or changed, then a different set of musicking characteristics would emerge. Within this structure, the robot has been embedded with a sense of esthetic that it can trust (believe to be true) and that the choices it makes are appropriate to it, co-creating inside the flow of musicking and unique to itself.
2) *Sound world*: The robot has a fixed sound library of roughly 1,000 short sounds, which were recorded through live improvisation, thereby embedding them with an essence of musicianship. These are triggered only when the wheels move. These are then either presented to the world in their raw state or treated in some way (time stretch, pitch shift, or both) using the Creative AI dataset as controlling parameters. The robot does not have the whole possible world of sounds, synthesis, and composition at its fingertips, but its sounds have a character and an esthetic basis which it can use to express its behavior and be unique to itself.
3) *Creative AI*: At the core of the creative AI, dataset is a world of embodied musicking captured through the EL process (described before) and through live interaction. These data are used to control every aspect of the AI, movement, sound production choice, and interaction goal. The dataset is also used to make choices about how the dataset is to be recalled and read by the algorithms (e.g., the read rate and ramp speed for each instance of wheel movement; discussed later). This means that the direct application of data into wheel movement and also the translations of that into sound object choice and therefore as music in the flow is imbued with the essence of embodied musicking that has been embedded in the core of the dataset. The version of the dataset in this application was a crude and small proof of concept. This has since been superseded by a larger project and a more comprehensive embodiment approach to the dataset.

But really these embedded belief structures are there so that the human musician can believe that the robot's behavior and responses are truly emanating through musicking, and to draw attention to that fact, this robot is a valuable co-creative presence inside a shared flow. We all know that this robot is really an assembly of plastic and metal components together with a couple of motors and a processor. But because the human musician can trust it believes in certain things and has been embedded with a certain notion of its world of concern through concepts such as affectual response, its range of sonic choices, and its behavior, the human musician can believe in it as a co-creative collaborator inside musicking, which in turn can lead to deep and meaningful human-centered interactions.

Technical Design

Hardware

The robot used in *EMRv1* was a Dexter *Go-Pi-Go* 1[2] with a Raspberry Pi3 model B[3] as the controlling computer. This system was used as the hardware was cheap, both the Dexter and the Pi had good online support and community forums, and there were plenty of ancillary peripherals available, such as cameras, which were equally cheap and supported. The *Go-Pi-Go* also came with an expanded version of the Raspian operating system (a Linux distribution) and included the libraries and dependencies to move the *Go-Pi-Go* completely with example scripts.

The robot was controlled remotely by the embodied AI in a black box system (discussed later) that broadcast movement parameters to the Pi. Onboard, the Pi was a simple script that received the transmitted parameters, translated them into wheel movement, and looked after the collision avoidance goal (below).

The embodied AI was built in *Max/MSP*[4] on a MacBook Pro and transmitted to the *Go-Pi-Go* using Open Sound Control (OSC)[5] protocols over wireless. *Max/MSP* is a graphical programming language for multimedia development. It is quick and simple to use and is specifically designed for real-time editing and interactivity. This made software development quick and simple and facilitated rapid prototyping. *Max/MSP* also organizes threading and concurrency internally.

EMRv1 Subsumption Architecture

The technical design of *EMRv1* was influenced by robotic subsumption architecture. This is a control architecture innovated by Rodney Brooks as an alternative to traditional AI, or GOFAI. Instead of guiding the robotic behavior by symbolic mental representations of the world, subsumption architecture "is a parallel and distributed computation formalism for connecting sensors to actuators in robots. A traditional way of describing these connections would be to say the subsumption architecture provides a way of writing intelligent control programs for mobile robots" (Brooks, 1986).

The subsumption architecture was designed to support multiple goals. These were (in order of priority) given as follows:

1) *Self-preservation*: The robot must avoid obstacles and not crash into the other musician or fall off the stage.
2) *Instinctual behavior*: If left alone, the robot would make music. This was driven by the Creative AI dataset (discussed before), which operated as its DNA of musicking creativity.
3) *Dynamic interaction*: The robot can, in certain conditions, be affected by the sound of the live musician. Using a process of simulated affect linking, the Creative AI could leap between

[2] *Go-Pi-Go*. Available at: https://www.dexterindustries.com/gopigo3/ (Accessed 2020/10/23).
[3] Raspberry Pi. Available at: https://www.raspberrypi.org/products/raspberry-pi-3-model-b/?resellerType=home (Accessed 2020/10/23).
[4] Max/MSP. Available at: https://cycling74.com/ (Accessed 2020/10/23).
[5] Open Sound Control. Available at: http://opensoundcontrol.org/introduction-osc (Accessed 2020/10/23).

related, abstracted, or unexpected datasets. Metaphorically, the robot's internal trains of thought would be triggered by phrasing (short-term temporal limits) and the dynamic impetus of the human.

A critical feature of the design of *EMR* is that each of these goals directly moves the wheels. It was essential that each goal is not part of an elaborate, logically flowing representation of a thought process, mimicking some kind of mind. As such, the overall design of the robotic system was modular, with each system directly accessing the wheels when in operation.

The overall modular design of the data flow is given as follows:

1) Live data sensors
2) Data wrangler
3) Affect mixing
4) Smoothing and deviation
5) Wheels move. Make sound

Although this may appear to be a linear flow based on a hierarchy, the subsumption design is embedded in module 2 *Data wrangler module*. With module 3 *Affect mixing*, enhancing the non-hierarchical approach. The code for *EMRv.1* is freely available as open-source on GitHub.[6]

The design of each module is as follows:

1) Live data sensors

This module coordinates and streams the live sensor data to various modules across the system. The live input consists of a line input signal from the collaborating musician (human or robot), and a stream of OSC data from the Kinect (x, y, and z coordinates of the head, body, left hand, and right hand). The audio from the line input was analyzed for dominant (fundamental) frequency and amplitude. These were then concatenated together as a series of lists and stored as dataset files for immediate use and access in later training sessions, and performances. The sample rate was flexible and was triggered by the incoming Kinect data. **Table 1** illustrates how these were saved as .csv files.

The operational processes of *1. live data sensors* module were given as follows:

i) capture the live sensor data and concatenate it into data lists;
ii) package the data lists as .csv files in the dataset local directory;
iii) stream each of the fields to other modules for use in real-time decision-making processes.

2) Data wrangler

This module generated the metaphorical trains of thought for *EMRv1* in the following two ways:

1) querying and reading from the files stored in the Creative AI dataset directory
2) generating outputs from four neural networks trained on the Creative AI dataset (discussed before).

The basic process for the querying and reading from the files stored in the Creative AI dataset directory was designed to symbolically represent the shifting nature of trains of thought as proposed by (Gelertner, 1994). The symbolic process was constructed as follows:

i) Choose a dataset file from the directory for a random duration (6–26 s)
ii) If an affect signal is received (see below), change the file immediately [goto 1]
iii) Choose a random line to start reading from the dataset file
iv) Start reading from this line for the random duration (3–13 s)
v) Read at a random procession rate (300–1,300 ms)
vi) Loop if triggered
vii) Parse and smooth all fields from the dataset as individual data atoms and send them to next module

The basic process for generating outputs from the four neural networks trained on the Creative AI dataset was triggered by the amplitude data received from three sources: 1) the live audio input (after FFT separation), 2) from the querying process before, and 3) from a short-term memory buffer that looped and recorded the live improvisation, and randomly read the audio from any point. Each of these was mixed and routed into each of the four neural networks, from which was generated x, y, and z data, which were streamed to the next module.

3) Affect mixing

This module received all the data streams from the dataset query, parsing process, and the neural networks and mixed them into the following two outputs: left wheel data and right wheel data. The mixing was controlled by a special process designed to symbolically represent affect and affect-linking (Gelertner, 1994) of a musician. In Vear 2019, I defined affect as "the mind's connecting response between sensorial input of external events with the internal perception of causation such as emotion or feeling, through time." This module translated this definition symbolically, the streams of amplitude data from the live input, the dataset parsing, and a randomly generated "drunk walk," would be used to trigger 1) local changes in the module such as mix and 2) global conditions such as dataset file selection. The basic process was given as follows:

i) randomly switch between input streams (1–4 s, or with a loud affect trigger)
ii) if amplitude is <40%, do nothing
iii) else if amplitude is between 41 and 80%, trigger a new mix (see below)
iv) else if amplitude is >80%, trigger condition changes across the architecture (new mix, new file read, restart reading rate, change smoothing rate, and change audio read in following modules)

[6]GitHub. Craig Vear. Available at: https://github.com/craigvear/Seven_Pleasures_of_Pris (Accessed 2020/10/23).

TABLE 1 | Example of the Creative AI dataset.

Id	Limb	X	Y	z	Freq	Amp
35	/Hand_Left	−0.31917	−0.295,487	1.376,182	161.538,467	0.322,659
36	/Hand_Right	−0.264,689	−0.213,074	1.28068	161.538,467	0.322,659
37	/Body	−0.397,107	0.106,659	1.222,754	161.538,467	0.322,659
38	/Head	−0.246,853	0.314,369	1.072035	161.538,467	0.166,799
39	/Hand_Left	−0.372,583	−0.077763	1.275,277	161.538,467	0.166,799
40	/Hand_Right	−0.256,269	−0.215,644	1.274,499	161.538,467	0.166,799
41	/Body	−0.387,607	0.108,567	1.23542	161.538,467	0.166,799
42	/Head	−0.239,018	0.316,554	1.083863	114.248,703	0.11613
43	/Hand_Left	−0.375,174	−0.039334	1.263,249	114.248,703	0.11613
44	/Hand_Right	−0.248,108	−0.212,755	1.270,422	114.248,703	0.11613
45	/Body	−0.365,129	0.119,221	1.260,812	114.248,703	0.11613
46	/Head	−0.23085	0.319,204	1.095646	31.987,429	0.131,989
47	/Hand_Left	−0.396,978	0.060928	1.210,986	31.987,429	0.131,989
48	/Hand_Right	−0.223,919	−0.181,981	1.253,668	31.987,429	0.131,989
49	/Body	−0.356,796	0.122,125	1.268,893	31.987,429	0.131,989
50	/Head	−0.227,154	0.319,129	1.099726	31.987,429	0.131,989
51	/Hand_Left	−0.456,557	0.253,543	1.086275	31.987,429	0.131,989
52	/Hand_Right	−0.208,468	−0.130,853	1.233,942	31.987,429	0.10922
53	/Body	−0.342,327	0.127,401	1.279,092	31.987,429	0.10922
54	/Head	−0.227,208	0.318,383	1.099777	31.987,429	0.10922
55	/Hand_Left	−0.323,183	−0.030516	1.199,504	31.987,429	0.10922
56	/Hand_Right	−0.173,354	−0.011939	1.149,591	31.987,429	0.10922
57	/Body	−0.332,003	0.131,065	1.283,651	31.987,429	0.10922

The mix function randomly selected which of the incoming data streams (x, y, and z from dataset read, x, y, and z, from live Kinect, x, y, and z from the neural network prediction) to be the output to the following module for the wheel movement. It was desirable that this involved multiple elements from these incoming streams being merged, metaphorically fusing different trains of thought into a single output.

4) Smoothing and deviation

The final stage in the dataflow process smoothed the output for each wheel using random slide properties of 15–450 ms. This would introduce a sense of push and pull in the final wheel response and sound generation, and like the other random processes were symbolic and metaphorical representations of rhythm and phrase generation. The last part of this process looked for deviations in changing data using a delta change function !(n - n-1). This was then sent to the wheel module.

5) Wheels move. Make sound

The left-wheel and right-wheel data outputs from the aforementioned module were rescaled and then sent via OSC and wireless to the *Go-Pi-Go* robot, which parsed them and moved the wheels. Simultaneously, these data were sent to the *Make Sound* module, which made independent sounds for each wheel. The data were rescaled between 0 and 1,177 so that it would trigger one of the minute samples held in its belief system (discussed before). These samples were then projected from speakers attached to the laptop.

DISCUSSION

The debate on whether intentionality is needed for creativity is still ongoing in the literature (Paul and Kaufman, 2014). *EMRv1* does not have a module in its subsumption architecture that deals with intentionality. Yet, I perceived moments of it intentionally responding to musicking, and also not responding to keys and triggers such as sonic impetus. This malleability in its response was intentional as I wanted to be surprised by what it did and did not respond to, in the same way that another human musician can choose to react to a musicking moment or not. This approach acts as a metaphor for my surprise when in musicking I make an unexpected response. In these moments (which happen regularly), I did not intend to respond, but something inside my being emerged. I cannot explain this, but I am aware of it, and the possibility of this is embedded in the experiential learning process and the Creative AI dataset with the symbolic AI making space for this to happen, or not.

It could be argued that this AI system is a passive passenger along the flow of musicking privileging the human musician as the central driver for all musicking decisions. And this is true, on a superficial level. All perceived interactions are from the human perspective, who in turn responds with a human-orientated decision. However, the embodied presence of the robot (in contrast with the presence of only a computer/non-anthropomorphic artificial system) did influence my human responses as I recognized its movements as being in the groove, due to them being based on my movements. This sense of familiarity with the movement (which in turn begat the sound) contributed to a sense that this *EMRv1* was inside its flow. This led to a sense of "meaning" as I felt that we were journeying together through a shared flow. Any points where I felt that the relationship was *concurrent*, *collaborative*, or *co-creative* further reinforced this, leading to a heightened sense of togetherness. I should add here that if *EMRv1* was left to perform a solo, it would do so without the need for

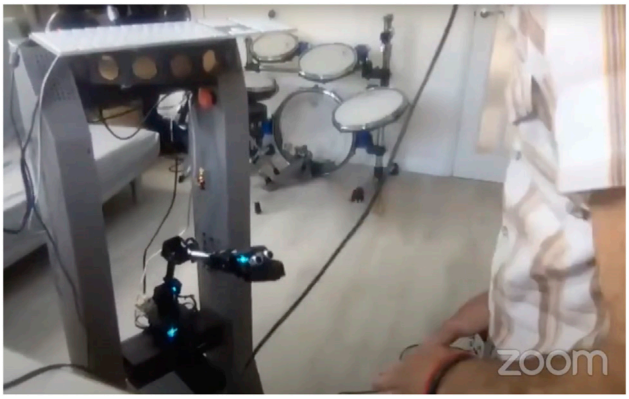


FIGURE 3 | A screengrab from an online performance of EMRv4 and the author as part of the Art-AI festival 2021.

human intervention. This was part of its "purpose of being." Similarly, when I placed two robots together, they performed a duet (Patabots, 2019a) (Patabots, 2019b).

It is a limiting factor here that due to lockdown and COVID pandemic restrictions, this research was unable to engage with other musicians and so remains anecdotal. But there is something in the way that *EMRv1* responds inside musicking that brought me closer to the improvisational relationships I have with other musicians. This is due to the goals and purpose embedded into the AI and robotic architecture of *EMRv1* being loose, and focused on surprise and novelty, as opposed to some elaborate mind-based model.

For me, this type of creativity happens inside a system that propagates principles of play and invention but is also bound by limits and parameters. Even the notion of free improvisation is bound to an individual's imagination and technique. Notions of "meaning" and "purpose" are therefore bound to enabling this system to operate within such parameters and limits. Meaning is the preserve of the human who recognizes that the system is in the flow, believes that its system is playing and inventing, and responds with creative playfulness. The robot AI has purpose, which is to play and invent within this system. Together, these create a system that can lead to emergent creativity. But this is not guaranteed; but neither is it guaranteed between human–human improvised musicking.

The consequence of this study is that it could signify a fruitful way forward of interpreting the concept of natural and artificial co-creativity. Considering playful creativity in AI as a defined system with a purpose, rather than a set of ingredients might unlock small-c creative projects. However, this also opens these applications to moments of failure as the system cannot be guaranteed to be creative all the time due to its inbuilt freedom, the integrity of the dataset, and the reliance of the human to comprehend what is understood as "meaning" in the flow.

CONCLUSION

Using the principles outlined earlier, the *EMRv1* project has created a co-creative system that responds to the interaction with a human musician through a cyclical relational process. It is important to note that the interaction with the musician begets movement as its primary goal for musicking and that this movement is embedded with the essence of embodied musicking because of the experiential learning process. Following this, the movement begets sound, which begets music such that all relationships between humans and AI are informed by phenomenon data captured within the embodied flow of music-making: either from the Creative AI dataset or through live interaction.

The subsumption architecture appeared to create a solution for an intelligent coping that followed the principles of the project (listed before). But due to COVID lockdown restrictions and budgetary factors, the testing of *EMRv1* was restricted to the author. However, these improvisations were presented on multiple occasions in front of the general public and peers, with encouraging responses and requests to try it out.

The design of *EMR* supported simple changes to its internal belief system that resulted in a change of behavior and esthetics. For example, swapping the source audio files for another set made the robot sound different. Changing some of its internal random parameters, especially in module 2 *Data wrangler* and module 4 *Affect mixing*, had a significant effect on its internal rhythmic and phrasing structures, thereby responding to the live improvisation with a different feel.

The ultimate goal of this research is *not* to find solutions to replace human creativity but to enhance it and move it forward into discoveries. In short, this research is seeking to find experiences like those emergent through DeepMind and Alpha Go's interaction with the professional Go players. In the 2019 film (AlphaGo, 2017), several of these professionals reflected that when they played with AlphaGo, they "see the world different [. . .] Maybe it's beautiful," and "like move 37, something beautiful occurred there"; "in a broad sense move 37 begat move 78 begat a new attitude, a new way of seeing the game he improved through this machine, his humanness was expanded after playing this inanimate creation" (AlphaGo, 2017).

I am hopeful, given the current trajectory and generation (v4) that this foundational work outlined in this article has proven to be a viable solution for such emergent creativity between musicking humans and robots. But as we are all still managing the COVID pandemic, and the limitation of face-to-face research, it may be a while before I am able to test EMR with another unbiased collaborator. However, as EMR grows with each iteration, the feeling of stimulated relationships in musicking grows. At a recent public talk/performance for the Art-AI festival 2021, I gave a demo performance of EMRv4[8] and I playing together (see **Figure 3**). It is interesting to note how the movement begets sounds and how the movement emits a sense of musicking, regardless of the sound produced [https://www.youtube.com/watch?v=LryGSo7MK74&t=3370s]. But, as mentioned before, this phenomenon needs a

[8]EMR v4 has now been migrated to Python and uses a hive of neural networks trained using the TensorFlow library, cooked using a dataset from the "embodied musicking dataset" repository. Available at: https://github.com/Creative-AI-Research-Group/embodiedMusickingDataset/tree/master/dataset.

considerable amount of further testing and validation and so remains only a mere hopeful conclusion.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: github.com/craigvear.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

AlphaGo (2017). *Film*. Director Greg Kohs.

Boden, M. (2003). *The Creative Mind: Myths and Mechanisms*. 2nd Edition. London: Routledge.

Brooks, R. (1986). A Robust Layered Control System for a mobile Robot. *IEEE J. Robot. Automat.* 2, 14–23. doi:10.1109/jra.1986.1087032

Brooks, R. (1991). Intelligence without Reason. Available at: https://people.csail.mit.edu/brooks/papers/AIM-1293.pdf (Accessed 202010 23).

Brooks, R. (1987). Intelligence without Representation. *Artif. intelligence* 47 (1), 139–159.

E. R. Miranda (Editor) (2021). *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity* (Springer).

E. Paul and S. Kaufman (Editors) (2014). *The Philosophy of Creativity: New Essays* (Oxford: Oxford University Press).

Gelertner, D. (1994). *The Muse in the Machine*. New York: Free Press.

Iacoboni, M. (2009). Imitation, Empathy and Mirror Neurons. *Annu. Rev. Psychol.* 60, 653–670. doi:10.1146/annurev.psych.60.110707.163604

J. McCormack and M. d'Inverno (Editors) (2012). *Computers and Creativity* (Springer).

McCormack, J., Gifford, T., Hutchings, P., and Llano, M. T. (2019). "In a Silent Way: Communication between AI and Improvising Musicians beyond Sound," in CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Springer), 1–38.

McCormack, J., Hutchings, P., Gifford, T., Yee-King, M., Llano, M. T., and D'inverno, M. (2020). Design Considerations for Real-Time Collaboration with Creative Artificial Intelligence. *Org. Sound* 25 (1), 41–52. doi:10.1017/s1355771819000451

Patabots (2019). *The Seven Pleasures of Pris*. Music Album. Available at: https://patabots.bandcamp.com/album/seven-pleasures-of-pris last (Accessed 202010 23).

Patabots (2019). *The Voight-Kamff Test*. Music Album. Available at: https://patabots.bandcamp.com/album/the-voight-kampff-test last (Accessed 202010 23).

Pearce, M. (2010). Boden and beyond: The Creative Mind and its Reception in the Academic Community. Available at: http://webprojects.eecs.qmul.ac.uk/marcusp/notes/boden.pdf (Accessed Feb 17, 2021).

Prosseda, R. (2014). THE ROBOT PIANIST: Educational Project of Musical Appreciation for Schools and Family Concerts. Available at: https://www.robertoprosseda.com/media/projects/000008/attachments/download/teotronico-en-2014.pdf (Accessed 202010 23).

Robocup (2015). Researchers Want Robots to Play in the World Cup by 2050. Available at: https://www.vice.com/en_us/article/nzep7q/robocup2015 (Accessed 202010 23).

Small, C. (1998). *Musicking*. Middletown: Wesleyan Press.

Thomsom, P., and Jaque, V. (2017). *Creativity and the Performing Artist: Behind the Mask (Explorations in Creativity Research)*. Cambridge, Massachusetts: Academic Press, 94.

Vear, C. (2019). *The Digital Score*. New York: Routledge.

Weinberg, G., Bretan, M., Hoffman, G., and Driscoll, S. (2020). *Robotic Musicianship: Embodied Artificial Creativity and Mechatronic Musical Expression*. Springer.

Zedan, B., and Pearce, T. (2017). "Ingold "Bringing Things to Life: Material Flux and Creative Entanglements"," in *State of Flux: Aesethetics of Fluid Materials*. Editors M. Finke and F. Weltzien (Berlin: Dietrich Reimer Verlag), 21–37.

Zedan, H., Cau, A., Buss, K., Westendorf, S., Hugill, A., and Thomas, S. (2008). "Mapping Human Creativity," in Proceedings of the 12th Serbian Mathematical Congress (Novi Sad: Springer), 1–14.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 26 November 2021
doi: 10.3389/frobt.2021.694177

Exploring Behavioral Creativity of a Proactive Robot

Sera Buyukgoz[1,2]*, *Amit Kumar Pandey*[3,4], *Marine Chamoux*[1] *and Mohamed Chetouani*[2]

[1]SoftBank Robotics Europe, Paris, France, [2]Institute for Intelligent Systems and Robotics, CNRS UMR 7222, Sorbonne University, Paris, France, [3]Socients AI and Robotics, Paris, France, [4]BeingAI Limited, Hong Kong, Hong Kong SAR, China

Creativity, in one sense, can be seen as an effort or action to bring novelty. Following this, we explore how a robot can be creative by bringing novelty in a human–robot interaction (HRI) scenario. Studies suggest that proactivity is closely linked with creativity. Proactivity can be defined as acting or interacting by anticipating future needs or actions. This study aims to explore the effect of proactive behavior and the relation of such behaviors to the two aspects of creativity: 1) the perceived creativity observed by the user in the robot's proactive behavior and 2) creativity of the user by assessing how creativity in HRI can be shaped or influenced by proactivity. We do so by conducting an experimental study, where the robot tries to support the user on the completion of the task regardless of the end result being novel or not and does so by exhibiting anticipatory proactive behaviors. In our study, the robot instantiates a set of verbal communications as proactive robot behavior. To our knowledge, the study is among the first to establish and investigate the relationship between creativity and proactivity in the HRI context, based on user studies. The initial results have indicated a relationship between observed proactivity, creativity, and task achievement. It also provides valuable pointers for further investigation in this domain.

Keywords: proactive robot, creative behavior, self-initiated behavior, human–robot interaction, social robot

OPEN ACCESS

Edited by:
Patricia Alves-Oliveira, University of Washington, United States

Reviewed by:
Filipa Correia, University of Lisbon, Portugal
Lawrence Kim, Stanford University, United States

****Correspondence:***
Sera Buyukgoz
sera.buyukgoz@softbankrobotics.com

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *12 April 2021*
Accepted: *14 October 2021*
Published: *26 November 2021*

Citation:
Buyukgoz S, Pandey AK, Chamoux M and Chetouani M (2021) Exploring Behavioral Creativity of a Proactive Robot. Front. Robot. AI 8:694177. doi: 10.3389/frobt.2021.694177

1 INTRODUCTION

Robots are becoming more and more a part of our lives. We encounter robots in our houses as assistants, at schools as tutors or peers, and at marketplaces as guides or shopping assistants. Robots appear not as tools but as social agents with a voice and a mind in our daily lives. Robots are producing behaviors that are intended to be supportive or helpful to the user. However, there is a need to investigate how such behaviors might be related to the users' expectations, causing some kind of confusion, or even related to the user's creativity. This will help in crafting the right level of behavior and suggestions that the robot should be providing. Robots can have different strategies to interact, such as *reactive*, when it acts only if there is a demand from the user, or *proactive*, where it acts even if there is no explicit request from the user. This study focuses on the proactive behavior of a robot with the intention to support the user in performing a task.

In organizational psychology, proactive behavior is defined as anticipatory, self-initiated, change-oriented, and future-focused behaviors (Grant and Ashford, 2008). Proactivity is described as a process of individuals influencing their environments (social, non-social, and physical) (Bateman and Crant, 1993) by intentionally taking initiatives (Bateman and Crant, 1999), by utilizing the combination of knowledge, perception, and ability to predict others' actions and consequences (Tomasello et al., 2005). Most human–robot interaction (HRI) studies are based on this definition from organizational psychology to define proactive robot behavior where robots must be anticipatory, self-initiated, and change-oriented toward future changes (Peng et al., 2019). In this study, we analyze the notion of proactive action as perceived from the user's perspective. Therefore, any act by the robot needs to be fulfilling the following two conditions for it to be perceived as proactive action (as opposed to reactive action) by the user: 1) there is an anticipation of the future situation. This can be either by a human controller or autonomously by the reasoning mechanism. 2) Based on the anticipation, if the robot is behaving without any explicit request from the user, it is self-initiated behavior of the robot from the user perspective. Again, such acts by the robot are instantiated either by a remote operator or autonomously. This situation is enough for us to perform our studies of proactive behavior from the user perspective. We are interested in understating the effect of such behaviors on the user, not how such behaviors should be created.

On the other hand, there is a notion of creativity, which refers to the novel product of value (Weisberg, 1993) or a person who expresses novel thoughts (Csikszentmihalyi, 2009). Being creative is the ability to change existing perspectives (Goncalo, 2019). In that sense, to be creative and proactive, both carry the similar notions of anticipatory, self-initiated, and future-driven behaviors. Therefore, in one sense, proactivity and creativity are highly coupled. Creativity, as the ability to produce novel ideas, is argued to be a necessity for proactive behaviors, and proactive personality is positively associated with creative behaviors (Joo and Bennett, 2018). In order to be creative, it is essential to have the ability to view things from different perspectives and generate new possibilities or alternatives in a unique way (Franken, 1994).

The majority of the examples we saw in robotics use interaction patterns to support users' creativity. Robots adopt the role of either a supportive agent that facilitates the user's creativity (Elgarf et al., 2021; Alves-Oliveira et al., 2019) or a creative peer that is collaborating with the user on a creative task (Law et al., 2019; Lin et al., 2020; Hu et al., 2021). In that sense, all of the examples put creative thinking of the user as an aim of the robot. Even some researchers claim that there is a positive effect of robot usage in education on a child's creative thinking (Ali et al., 2019).

In reality, life is not focused on creative thinking, even to the extent that educationalists complain that the current education system is blocking creative thinking. Ken Robinson stated in his TED talk that school kills creativity. The current education system depends on convergent thinking, asking for the answer to a question, rather than divergent thinking, asking how to reach that answer (Ritter et al., 2020). If we think about the task-based robotics system, which is popular in robotics systems, how could they cope with supporting the user's creativity?

In our study, on the proactive side, we focus on the robot being proactive toward the user who is completing the task of cooking recipes. The robot exhibits proactive actions by predicting what users try to achieve without the user asking for information or support and instantiates a set of verbal communications with the user. On the creativity side, we focus on the two aspects of creativity: creativeness observed by the user in the robot proactive behavior and the creativity of the user while leading toward a task of reaching a cooking recipe.

Our motivation is to understand the effect of proactive behavior and the relation of such behaviors to the two aspects of creativity discussed above. This study aims to present the results of a set of pilot studies of different behaviors of the robot. The idea is to understand and explore the limitations and pointers for conducting a full-scale research project. To our knowledge, it is the first study of its kind, which is trying to explore the connections between observed proactivity, creativity, and task achievement, in a setup of users with mixed backgrounds. The initial results have indicated some interesting relations among various attributes. At the same time, it is hinted that it would be too early to draw a definite guideline and conclusive relations about the optimum behaviors of the robot. Nevertheless, our findings suggest various parameters, which need further investigation when such social robots will serve people in day-to-day activities, where a series of actions are needed.

2 BACKGROUND

Creative thinking is defined as a skill that produces novel and valuable ideas (Sternberg, 2010). It is a way to consider things from a different perspective, be creative, and have a different look at daily life problems. It depends on the knowledge of the individuals. There is no prior way to define creativity. Although the creative contributions are classified under eight headings (Sternberg, 2010), evaluating the amount of creativity is not clearly defined. Some types could have greater amounts of novelty than others. Creative thinking is not the equivalent of divergent thinking. However, divergent thinking tests can be used for estimating creative thinking (Runco, 1993). Divergent thinking is defined as the ability to produce diverse ideas (Runco, 1993).

Producing creative ideas is affected by the environment. However, creativity requires a moderate level of focus. Therefore, when a person is interrupted or the attention process gets disturbed, it might affect the creativity performance (Woodman et al., 1993; Wang et al., 2014). In that sense, potential interruptions by the robot to the human performing a task, which might occur because of the proactive interactions, might have effects on the creativity of the person. Furthermore, interruptions might also come from the environment and/or other agents during the interaction.

Several human–human interaction studies are providing an interesting understanding of the relationship between receiving interruptions and being creative. Different types of moods on the interruptions are highly studied in the cognitive science domain—most of them are based on variations of the tone of the interruptions. Studies have shown the positive effect of a positive mood of the interruptions on creativity (Baas et al., 2008; De Dreu et al., 2008) to generate new ideas. Another study shows the strong relationship of receiving different types of interactions and being creative. Studies have shown the positive effect of positive feedback on creativity to generate new ideas (George and Zhou, 2007; Gong and Zhang, 2017). The positive interruptions carry combinations of positive feedback—validating or praising the user—and constructive feedback—question users' actions and lead them to think about the solution they find (Gong and Zhang, 2017). The effect of negative feedback in connection with the frequency of positive feedback is also studied. For example, if the level of positive feedback is high, then negative feedback positively affects the creativity performance (George and Zhou, 2007; Gong and Zhang, 2017). Different studies are highlighting the effects of interruptions' tone, type, and frequency to be creative.

2.1 Proactivity in Robotics

The previous aggregated definitions of proactivity align with many of the recent studies in robotics for defining a robot's proactivity as the anticipatory action initiated by the robot to impact itself or others (Peng et al., 2019). It is defined as acting before it is requested (Ujjwal and Chodorowski, 2019). The proactivity of robots is studied in different implementation methods by using different initiations. The most related ones are as follows: 1) *anticipating user needs*, which is when the robot understands the user's needs and offers its support (by acting or interacting) for clarifying confusion (Pandey et al., 2013) or for providing suggestions (Peng et al., 2019; Baraglia et al., 2016; Grosinger et al., 2016; Myers and Yorke-Smith, 2007; Bader et al., 2013; Zhang et al., 2015; Ujjwal and Chodorowski, 2019), 2) *for anticipating* possible plan failure and plan repair (White et al., 2017), 3) *for preventing* future hazards (Bremner et al., 2019), 4) *improving the robot's knowledge & seeking information*, such as the robot asking the user for validation or the robot asking the user to identify gaps in the robot's knowledge (Lemaignan et al., 2010) (Moulin-Frier et al., 2017), 5) *seeking engagement and interaction,* such as proactively seeking the user for interaction (Garrell et al., 2013) or continuing interaction (Liu et al., 2018), 6) *adapting to the user*, such as the user's action while working together (Awais and Henrich, 2012), following the speed of the robot while considering constraints that the user needs to meet (Fiore et al., 2015), or considering the user's habits and arranging the robot's action not to become annoying (Rivoire and Lim, 2016) or enacting humanlike behaviors while reaching the object (Cramer et al., 2009; Han and Yanco, 2019), and 7) *adapting robot roles*, such as changing the robot from the leader to the follower during cooperative manipulation tasks (Thobbi et al., 2011; Bussy et al., 2012).

There are various modalities to exhibit proactive behaviors, from manipulation to verbal suggestions. However, while considering the limitations of social robots, it is already challenging to manipulate the shared environment physically since social robots' manipulation capabilities are not as precise as those of collaborative robots or industrial lightweight robots. On top of that, some HRI studies show that humans are more accepting of the robot's proactivity when the robot includes the human in the decision instead of the robot applying the decision itself (Kraus et al., 2020). Therefore, to simplify the complexity of the experiment and to avoid imposing a decision, we chose to instantiate communicative actions, thus limiting other modalities to make the robot proactive.

2.2 Creativity in Robotics

Creativity is studied in different areas of robotics. The majority of the task definitions are inspired by figural and verbal creativity. For instance, users are assigned to tell a story (Elgarf et al., 2021) or draw a figure (Alves-Oliveira et al., 2019). Robots produce behaviors for supporting the user's creativity. Robots present a collaborative behavior where they study social aspects and engagements of robots with turn-taking principles. The robot's role depends on supporting the user's creative behavior, where the robot is most likely asking the user to think about their decision and lead them to be creative. As in the example of Kahn's Zen Garden (Kahn et al., 2016), the robot follows a pattern of interaction to foster the user's creativity.

The robot's creativity is also explored in various studies. For example, the study of human–robot collaborative design (Law et al., 2019) aims to facilitate creativity of the user and be creative as a robot. Both the robot and the human are playing creative roles in the task. So, the robot supports the user and tries to be creative in the decision process to design a pattern. In their study, the robot's and the human's creativity share the same definition of being unexpected, novel product creation. Meanwhile, in the study of co-creativity (Lin et al., 2020), the focus is on facilitating the robot's creativity by getting feedback from the user while collaborating to draw a figure. The main point is to lead the robot to more creative outcomes (Hu et al., 2021), since the robot mostly creates ideas during collaboration.

The creativity process requires an environment of help or individual effort to develop ideas for self-initiated projects (Apiola et al., 2010). In this sense, a robot's creativity depends upon the robot's ability to produce helpful information. In this study, the creativity of the robot focuses on the definition of creating useful help. This created help is communicated to the user without being asked for it and is hence used as a proactive robot behavior. As such proactive behaviors are interrupting the users, they have a potential effect on the creativity of the user as well.

Thus, this study explores the behavioral aspects of creativity, inspired by our previous study of creativity and proactivity (Buyukgoz et al., 2020), which hinted about the possible relationship between these two aspects, and also suggested the need to investigate the duality of creativity: *creativity of the robot* and *creativity through the robot*. More specifically, in that work, we developed a study to experiment with the robot's proactive behavior when there is no task explicitly assigned. The behavior occurs as a set of verbal interruptions toward the users as a result of anticipation of the situation. The study was for understanding how the robot's proactive behavior is perceived as a creativity of the robot and its effects on the user's creativity in the HRI scenarios. Although, as discussed earlier, the frequency of interruptions also has a role to play in creativity, we also instantiated different levels of proactive interactions to explore the effects.

3 DESIGN OVERVIEW

This study explores the behavioral aspect of creativity during HRI using a robot as a proactive agent, creatively engaging in a

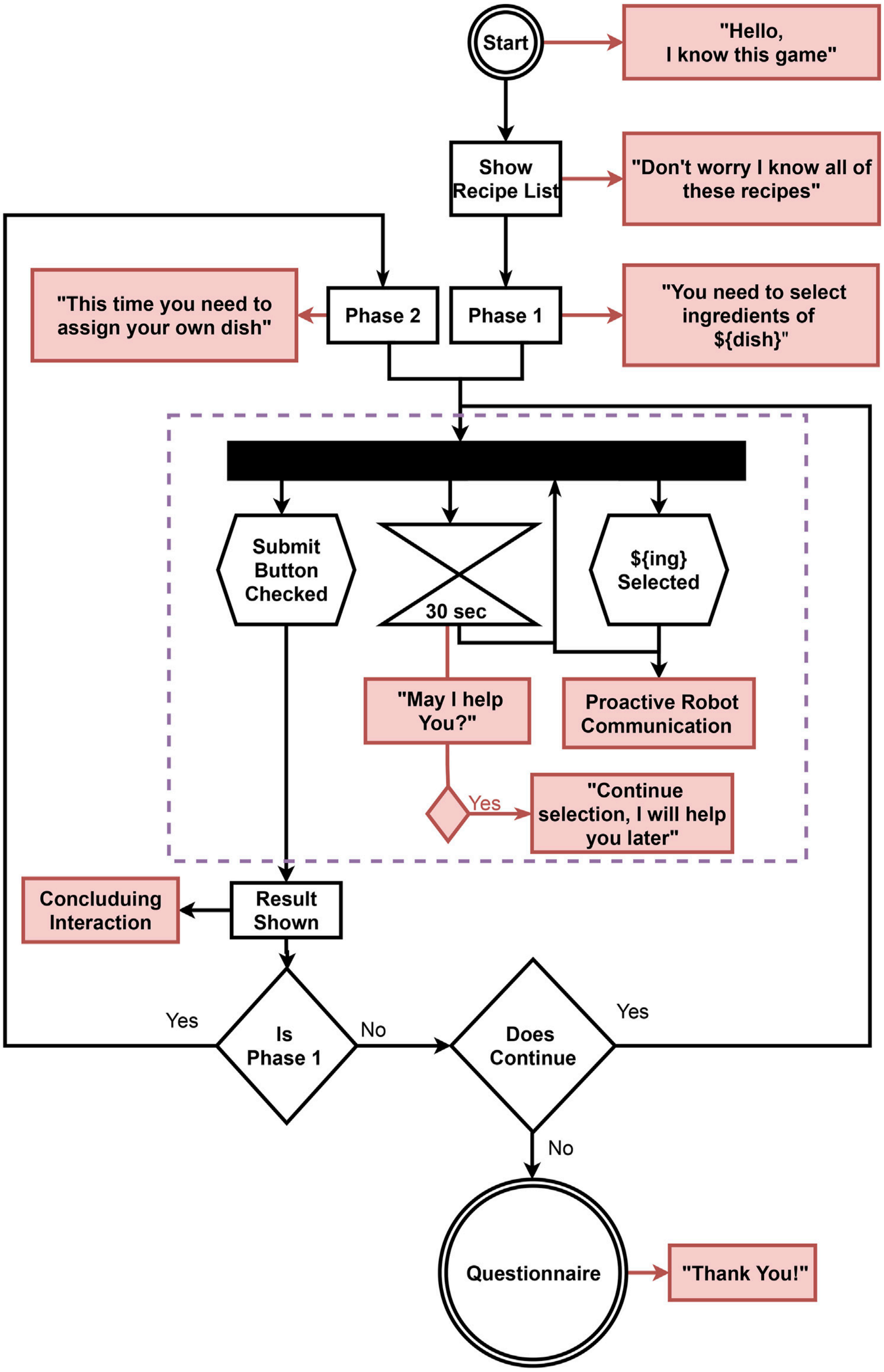


FIGURE 1 | Activity diagram of interaction flow indicates the actions of the task, the robot, and the user. Black rectangle boxes represent the pages of the task and diamond boxes show the decisions of the task for general flow of phases. Robot responses are indicated with red boxes. Red boxes with black arrows represent the sentence that will be created. The user's actions are represented with hexagon boxes.

task performed by the user and generating suggestions as communicative behavior to guide the user while achieving their intended goal. In this regard, *proactive behavior* of the robot is defined as instantiating behavior for suggesting the users, where the robot's intervention is not necessary or not requested by the user. We have considered a common home scenario task—cooking a recipe—that is explained below.

3.1 Task Definition

We have chosen a task in which the user could be creative and would not necessarily need the robot's help to complete the given assignment. Generating cooking recipes can be seen as a creative task, in which the users could converge toward different recipes by using a similar set of ingredients.

Most likely, users have differing knowledge about the typical recipe. That is why users were provided with seven recipes with the dish's name and ingredients to set the expected ground for the cooking recipe scenario. Each dish has six different ingredients. The recipes are different. However, each dish shares one or more ingredients with other dishes. In total, twenty-two ingredients were given in alphabetical order from different categories: proteins; chicken, foie-gras, ham, lardon. Vegetables; garlic, pumpkin, spinach, truffle, onion, pepper, nutmeg. Dairy products; milk, cheese, butter, eggs. Processed stuff; bread, barbecue sauce, stock, wine. Sugary stuff; sugar, honey. Basics; flour. Typical French recipes such as quiches, soups, and toasts were chosen to eliminate the hassle of learning new recipes. The whole task is divided into two phases to first study the effect of proactive behavior for a predefined recipe and then give the user a chance to be creative by eliminating the predefined recipe. With each user, the experiment starts with Phase 1 and continues with Phase 2.

Phase 1: a dish is assigned to the user. The participant should select the exact ingredients for the given recipe. Thus, the participant and the robot both know which recipe is targeted.
Phase 2: the user is asked to create a dish by using the given ingredients. The robot does not know about the target dish.

3.2 Design of the Proactive Behavior

In this study, the proactive behavior of the robot uses shared principles of creativity and proactivity. Those principles consist of 1) being anticipatory, based on a particular state, 2) self-initiated, producing proactive suggestions without them being explicitly demanded by the user, and 3) future-driven, trying to converge toward the needs of the goal. A rule-based system is developed for the robot to instantiate verbal suggestions depending on the user's task. Rules are selected based on the task's needs and the understanding of intention recognition. With the help of the rules, reasoning occurs to instantiate the parameters of the proactive suggestions. The robot's knowledge and this reasoning result are proactive robot communication (see **Figure 1**). The decision flow of instantiating the proactive robot communication (aka proactive suggestion) is shown in **Figure 2**.

The set of rules varies according to the need of the task. The tone of the proactive behavior slightly differs depending upon whether there is a target dish assigned by the system (Phase 1) or not (Phase 2). The decision-making processes to instantiate the proactive interaction, for Phase 1 and Phase 2, are shown in **Figures 2A,B**, respectively. In both cases, instantiating the proactive robot communication starts with intention recognition of the user actions. Intention recognition is the recognition of the user's target dish by interpreting the robot's knowledge of the dishes and the ingredients that the user has selected so far. The recognition process is a simple rule-based mechanism that checks how close the user is to achieving one goal. The user's intention is based on the least number of ingredients left from the set of known dishes. The intention is either the list of dishes or a single dish, depending on the situation of the selected ingredients so far. The user is also free to move away from the set of dishes that the robot knows and create their own dish by selecting a new list of ingredients. The user is assumed to be reliable and collecting ingredients to complete a dish. The user willingly performing a faulty behavior to deviate the intention recognition is not handled in this recognition mechanism. In the beginning, the intention of the user is all the dishes that the robot knows. Then, the recognition mechanism updates the user's intention for every change in the state (adding or removing an ingredient). Respectively, the system initiates the new proactive suggestion.

Different sets of rules are used to instantiate the sentences' templates, depending on whether the goal is assigned (phase 1) or not (phase 2). In phase 1 (see **Figure 2A**), it is crucial to accomplish the assigned dish by selecting the exact ingredients of the target dish. That is why intention recognition responds to each change in the state by updating the list of intentions. Updating the intentions triggers the process of instantiating the sentences' templates. The next step of "*Goal in Intent*" (as shown in **Figure 2A**) is to check if the targeted dish (which is the goal as shown in **Figure 2A**) is part of the intention list or not. This reasoning gives the impression that the user is on the right track. Then, the length of the intention list is checked to elaborate more on whether the user follows one specific dish or there are still multiple possibilities. For the cases in which the goal is in the intention list, the robot gives feedback type of suggestions that give the information about the status of the action. The action represents the selected ingredient and is denoted by $< a >$. The action status could be *True* or *False* depending on whether the played action complies with the goal's recipe. For example, say *Fois Gras Toast* is assigned as a target dish, and the user has already collected *foie-gras* and *truffle*. The recognized intention is *Fois Gras Toast*. Now the user collects *butter*: this action is *False* because collecting *butter* does not comply with *Fois Gras Toast*'s recipe since the recipe does not include butter. Therefore, the instantiated interaction will look like "*You lost a bit. You should remove butter.*" Here, it is interesting to note that such feedback was not requested by the user. Therefore, from the user's perspective, it is a proactive action, as the robot is acting by itself by anticipating the future situation.

In phase 2 (see **Figure 2B**), it is crucial to keep up with the user to assist the user in accomplishing the user's goal. The difference from phase 1 is that the robot is unaware of the goal: the user chooses it. The robot uses intent recognition to predict the goal of

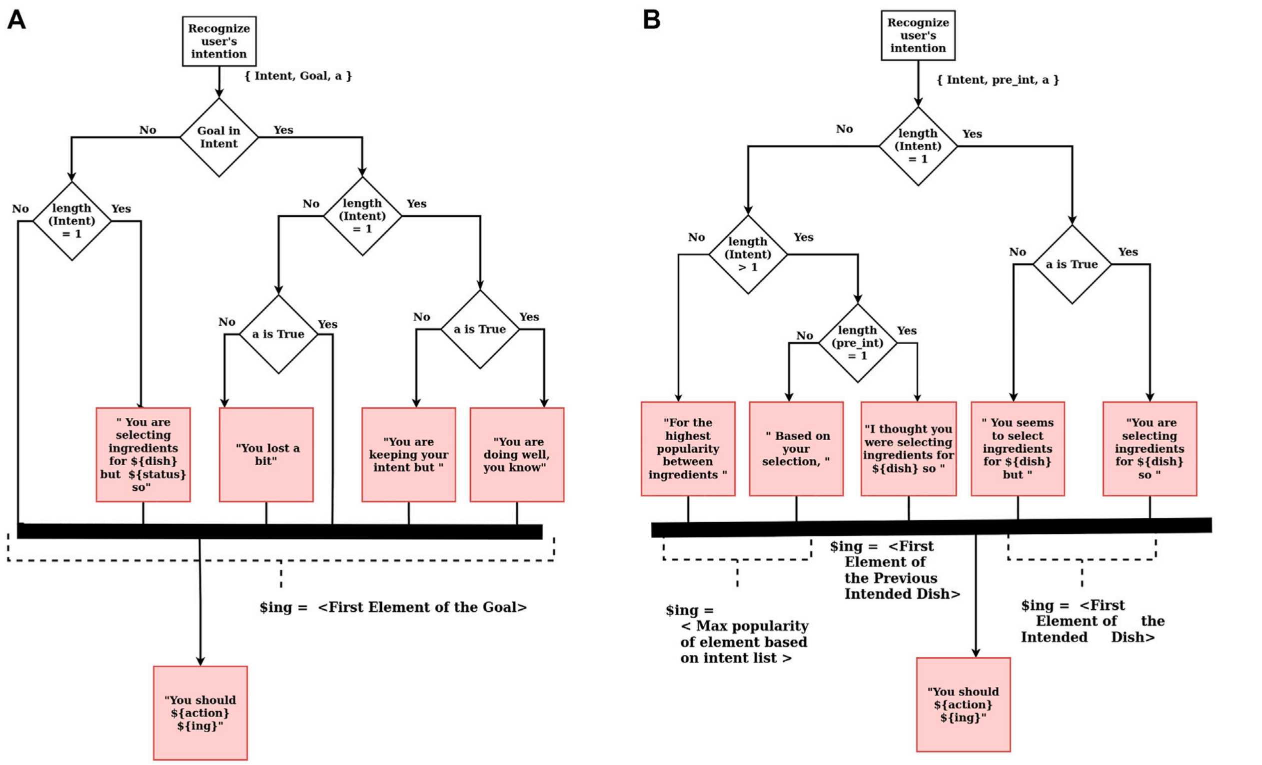


FIGURE 2 | Diagram of proactive robot communicative is showing how proactive suggestions are initiated. It is starting with recognizing the user's intention. The list of the user's intent passes to the reasoning mechanism which is based on layers of rules. Red boxes show the robot response. In boxes, the template is given to create a sentence. Between curly brackets of ${. . .} filled with symbolic representation of action (${action}), object–ingredient (${ing}) or dish (${dish})–or the result of system (${status}). Each box concatenates to each other on the road to create proactive behavior of the robot.

the user. The rules of the proactive suggestion focus more on the user's consistency than on assisting. That is why the current intention list (which is the intent as shown in **Figure 2B**) and the previous intention list (which is the *pre$_i$nt* as shown in **Figure 2B**) are used for reasoning. After updating the intention list, it is checked whether or not the user's intention is a specific dish. This means the length of the intention list is equal to one; therefore, a single intent is recognized. This case is treated similarly to a supposed target goal. That is why the status of the action is checked, as explained for the similar situation of Phase 1. If there is no specific intent, the system tries to lead the user by suggestions. The reasoning about suggestions starts with checking if there is any intent in the intention list or not. If the length of the intention list is equal to zero, the system tries to lead the user by suggesting the most frequent ingredient. If there is an intention which means the length of the intention list is non-zero, the length of previous intent is checked to be equal to one to determine if the user had a goal. In that case, the suggestion instantiates for explaining its reasoning and objectives of the previous goal. For example, in this situation, the robot said, "*I thought you were selecting ingredients for Fois Gras Toast so you should select truffle.*" Otherwise, the suggestion relates to the most popular element in the list.

3.3 Implementation Details

The Pepper humanoid robot [description can be found in the work of Pandey and Gelin (2018)] interacted with the participants during the experiment. The robot followed the actions of the participants from a web-based interface and instantiated interactions from the Android application of the robot. The task was presented on the laptop with a web-based interface. The participant can only take actions and decisions with the laptop. The graphical user interface (GUI) that the participants faced is shown in **Figure 3**. The connection between the robot application and the web application is made using a Firebase database (Moroney and Moroney, 2017).

The interaction flow is shown in **Figure 1**. The diagram shows the combination of web-based task flow, the robot interruptions, and the participant's actions. The task and behavior system are separated from each other to divert the participants' focus from the robot to the task.

4 EVALUATION

The in-person experiment was designed and conducted at the SoftBank Robotics Europe facility.

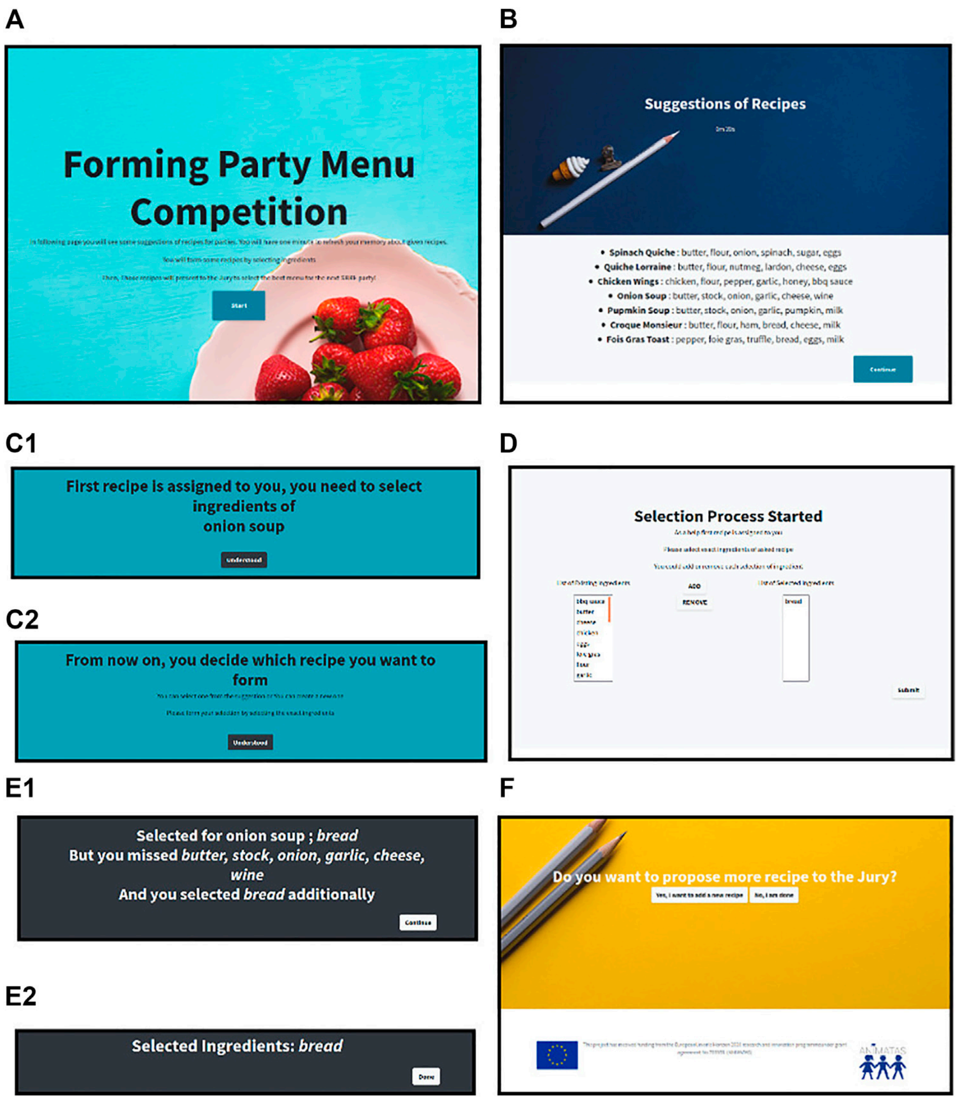


FIGURE 3 | GUI of cooking recipe task; diagram combines the pages of the web-based task application. After the robot connected—additional pages omitted, **(A)** is welcoming the user and explaining about the aim of the task. **(B)** is only visible for 1 min, and it presents the examples of recipes. **(C)** is changed according to phases of the task, (C1) is phase 1, where a dish is assigned, and (C2) is phase 2, where the participant has freedom of choice. **(D)** is the page where phases of the task occur, where the robot proactive behavior is activated. It is identical for all phases. **(E)** is the result pages after each selection process. (E1) has a specialized view for phase 1 which provides more information. (E2) is the result page for phase 2 which only shows the recent information of selection. **(F)** is the page that comes after phase 2 and lets the participants continue to create more dishes or finalize the task.

4.1 Participants

A total of 30 participants (11 female and 19 male, average age 32.23, standard deviation 6.76) participated in this experiment. All of them were employees of SoftBank Robotics Europe, Paris. They had some experience with the Pepper robot. However, they had different backgrounds: technical (hardware and software) and non-technical (marketing, communication, and welcome desk). The participants were also fluent in the language of the experiment: English. All participants gave their consent and signed a form giving permission to use and share their anonymous data for scientific purposes.

4.2 Hypotheses

We aim to study how the robot's creativity (which is instantiated through the proactive interaction) affects 1) the perceived creativity of the robot and 2) the creative process of the human during an HRI task scenario. Recall that creativity is seen as bringing novelty, and proactivity is anticipatory behavior aiming to help in the task. Therefore, we developed the following hypotheses to study their relation and effects on the user's perception.

H1: Proactivity and perceived creativity of the robot; the proactivity of the robot behavior will affect the perceived creativity. Proactivity in the robot behavior and its perceived creativity are related.

H2: Proactivity and the user's creativity; there exists a link between the robot's proactivity and the resulting creativity in the user (measured by the novelty of the products that the user creates in the HRI task). That is, proactivity of the robot and the facilitated creativity in the user are potentially related.

H3: Proactivity and goal achievement; there is a relationship between the robot's proactivity and the success of the HRI task. That is, the proactive behavior of the robot can help to achieve the goal of the task.

H4: Proactivity level and user perception; different levels of proactivity of the robot will have different user experiences on the perceived attributes, including perceived and facilitated creativity.

4.3 Study Design

A between-subjects study was conducted with one independent variable, the proactive behavior of the robot, which has three conditions: *high*, *medium*, and *no proactive*. The different conditions of proactive behavior aim to change the frequency of exhibiting proactive interactions. Under full proactive conditions, it is expected that the robot will provide feedback after each action of the user. On the other hand, under no proactive condition, the robot is not providing any feedback. An intermediate condition (medium proactivity) is detailed below, along with details of the other conditions. The robot also talks at the start and between each phase of the task. Participants were randomly assigned to different conditions.

4.4 Conditions

The robot followed the general flow of the interaction with the participant, as shown in **Figure 1**. The main aim of the added interactive behavior is to balance the frequency of the robot's talk between different conditions. In a between-subject study, the participants only interacted with one of the conditions. Three different conditions of the robot's interactive behaviors are instantiated for this experiment. These conditions are as follows:

Condition 1: no proactive behavior. The robot does not provide any explicit or implicit directions to the user in terms of the status of the action. After each step of the ingredient selection process, the robot simply utters, "oh[1]." We choose "oh" because it is a "knowledge state marker," that is, when "oh" is uttered by the robot, it informs the user that the action they have undertaken is understood by the robot but does not give (either positive or negative) feedback on this action. Thus, as a knowledge state marker [as described in Heritage (1984)], "oh" is used as a neutral token to acknowledge to the user that their action has occurred.

Condition 2: medium proactive behavior. Under this condition of the experiment, the robot provides communicative proactive action at every third action of the participants and utters "oh" in other steps. The frequency of interventions is decided based on the approximate number of actions played in each phase. If everything goes well, the participants need to play six actions to accomplish the goal. It is decided that the robot acts at least every third action to (at minimum) have support at half of each phase. The proactive actions are instantiated through a response trigger mechanism described in **Section 3.2**.

Condition 3: high-level (full) proactive behavior. Under this condition, the robot instantiates and provides communicative proactive action after each action of the participants.

Thus, the kind of information the robot provides under medium and high proactive conditions is related to the ingredients selected by the user for a dish. At the end of each phase of the task, which is supposed to result in a recipe, the robot provides a summary of the selections. The response of the robot is instantiated by a matching mechanism using the database of known recipes, their ingredients, and the selection of ingredients by the user. If the participant created a new recipe (mainly by selecting a novel set of ingredients) that the robot could not find a matching recipe for, the robot asked the name of the potentially "new" recipe of selected ingredients to use this information for interaction purpose.

4.5 Setup

The experiment was conducted in the various meeting rooms of SoftBank Robotics Europe. The experimental setup is shown in **Figure 4**. The participant sat in front of a laptop to get engaged in the task. A Pepper robot was placed relatively to the left or right of the user. Participants manipulate the task environment on the screen of the laptop through a mouse or track pad. A self-report questionnaire is attached to the task and automatically pops up once the task is over. As a part of COVID 19 guidelines, all equipment was sanitized before and after each session. Participants were left alone in the room with the robot during the study.

4.6 Procedure

Procedures for all conditions are identical except the robot's interruption frequency during the execution of the phases of the task. After signing the informed consent form, participants are informed about the experiment. Participants were given the choice of suggesting as many recipes as they wished for the upcoming hypothetical company event. The way to suggest is by using the online platform. They were informed that the online tool would guide them on how to proceed. Sample recipes were given to remind them how ingredients may be used. They could list as many recipes as they wanted while the Pepper robot accompanies them. They were reminded to be aware of the existence of the Pepper robot. Then, the experimenter left the room. Each participant interacted with one condition of the proactive robot behavior (condition 1: no proactive, condition 2: medium proactive, and condition 3: high proactive), which was assigned randomly and maintained during both the phases of the task. As a result, each participant can generally work on two kinds

[1]In the implementation using a Pepper robot, the exact token used was "oo," as "oh" sounded unnatural given Pepper's text-to-speech component.

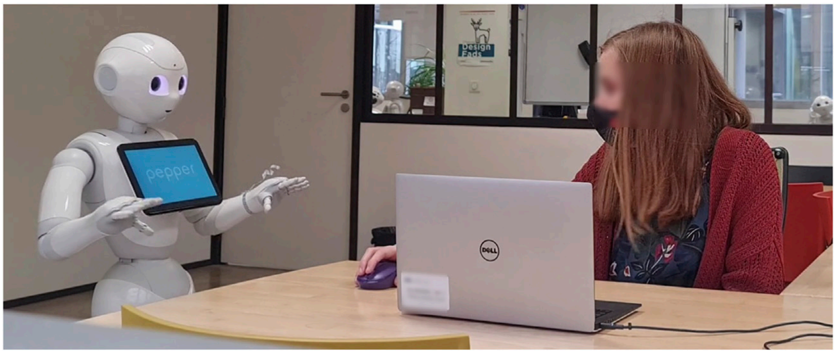


FIGURE 4 | Set up of the experiment. Participant sits in front of a laptop and the robot is placed next to the user. They share the space as the robot is looking at the screen over the participants.

of dishes: one that is assigned to them and one that they created. Participants were also allowed to proceed without selecting any ingredient by submitting the result without collecting any ingredients at the execution of the phases in page (in **Figure 3D**). The proactive robot behavior is initiated depending on the robot's knowledge. In this experiment, the task space and the participant's action in the task space were used to enrich the knowledge of the robot. The robot stayed ignorant of the other possible actions from the participant or the shared environment. After each participant had completed the task, the self-report questionnaire was submitted. The self-report questionnaire is attached to the task interface. It automatically pops up when the task has been completed. After the participants completed the task and the self-report questionnaire, the experimenter came back to the room for a small interview.

4.7 Measurement

Different evaluation metrics are used to investigate different aspects of proactivity and creativity. Our measures are divided into three sections to assess the following:

Creativity of the user; to define and evaluate the participants' creativity, metrics were inspired by divergent thinking. Thus, the creative thinking of the user often links with divergent thinking tests. Traditional methods of scoring divergent thinking (i.e., fluency, originality, and flexibility) are the most used methods for assessing the potential of creativity (Runco, 1992). In this study, we created an assessment influenced by the Torrance Test of Creative Thinking (Torrance, 1974) by focusing on fluency in the task—*How many dishes were achieved?*—and originality—*How many new dishes were created?*. These two scores are used to measure the creative thinking of the user. The total number of dishes is summed at the end of each task. It included phases 1 and 2 and repetitions of phase 2. The number of new dishes is the count of all dishes created in phase 2 and repetitions of phase 2. Dishes in the list of ingredients that have the same recipe as dishes in the recipe list are extracted.

Creativity of the robot; to assess some creative aspects of the robot's behavior, a different self-report questionnaire was used to assess the participants' perception of the robot. The questionnaire is a combination of different sections to assess demographic information, participants' personality and creativity using a Likert Scale, acceptance of social robots from the ALMERE questionnaire (Heerink et al., 2010), comprehensive impression of user experience from the User Experience Questionnaire (UEQ) (Schrepp and Thomaschewski, 2019), and some specific questions directly related to engagement, proactivity, task, and overall interaction. In this study, we did not include all the scales from the questionnaire, such as ALMERE and UEQ. Instead, we included the scales that could be applicable to the defined situation such as perceived adaptivity, perceived enjoyment, attitude, perceived usefulness, trust, and dependability. The scales assess the participants' perception about the robot's creativity on generating proactive actions that are task-oriented.

Effect of proactivity; to assess the effects of different conditions of proactive behavior on the task, we check the success rate of phase 1. In phase 1, a random dish is assigned to the participants, and it is expected of the participants to select the exact ingredients which were shown to them earlier. We also analyzed the time that they spent during the selection process in phase 1. The spent time is calculated by the time that passed since the user started to select ingredients until they submitted their selection by clicking the submit button.

5 RESULTS AND DISCUSSIONS

This section presented an evaluation for our four hypotheses (outlined in **Section 4.2**). We take the user study from 30 participants, 10 for each condition (no, medium, and high proactive behavior of the robot). We conducted a one-way analysis of variance (ANOVA) in each of our hypotheses to see if they are different for the three conditions: no (n = 10), medium (n = 10), and high (n = 10) proactive behavior with a *post hoc t*-test to compare differences in paired conditions. To test our hypotheses, recall that we use a combination of qualitative and quantitative measures (as is further illustrated in the following sections). For the qualitative data, we analyze the results of the questionnaire as the post test given to the participants. For

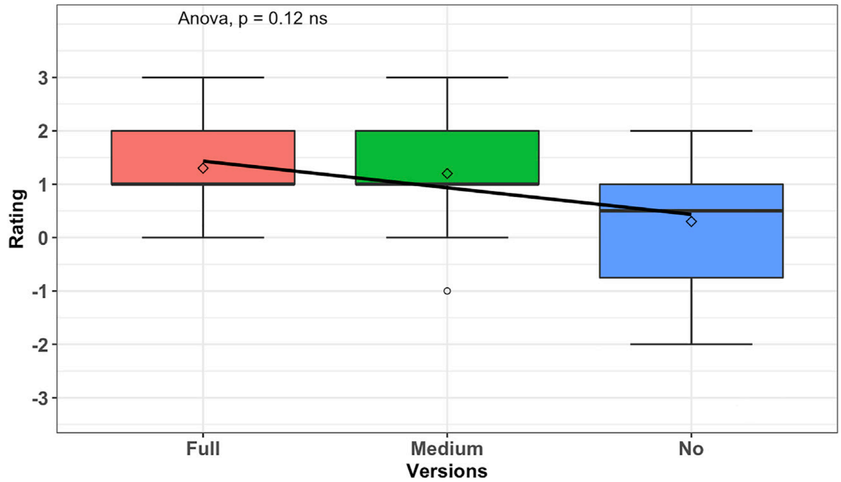


FIGURE 5 | Analysis result of novelty from UEQ; the graph is scaled on positive as creative and negative as dull. It shows that in each version of proactive behavior, the robot rated as creative; however, there is a visible difference on the mean from full to no proactive behavior.

quantitative data, we use the meta data generated from the results of the task (such as the number of times a dish was created). The data are presented as the mean and the standard deviation.

Before conducting the ANOVA test, we check that the following assumptions are not violated: 1) no significant outliers, 2) test for normality (by Shapiro–Wilk's test), and 3) homogeneity of variances[2] (by Levene's test). We do not check for the independence of observations, as each participant belongs only to one condition/group. For the test to detect outliers, we check outliers for our quantitative data collected, but we keep the data since we manage to figure out the reason for the outlier. That is discussed in the following section. However, the qualitative data outliers are subjective reports and are essential for our analyses (e.g., how was the perceived adaptivity of the robot?). However, we do report this range of differences in the user's opinion. The results are shown according to each hypothesis in the following sections. The significance of p is denoted by stars *, from high impact (***$p < 0.001$) to low impact (*$p < 0.05$), and nonsignificance is denoted by (ns $p > 0.05$). Qualitative observations are discussed in the following section, along with various attributes and pointers for further investigation.

5.1 Proactivity and Creativity

The *Novelty* scales of UEQ (User Experience Questionnaire) measures how much the design of the robot's behavior is perceived as creative. The user is asked to rate from dull to creative, with the statement "*In my opinion, the idea behind the robot's behavior and design is –*" on a 7-point scale (-3: dull to 3: creative). We conduct a one-way ANOVA test to see if the perceived creativity of the robot was different in our 3 group conditions. We run tests before conducting the one-way ANOVA test on the dependent variable (creativity) to check that the assumptions are met. We only have one outlier which is not extreme; one user in the medium condition rated the perceived creativity of the robot as dull (-1) compared to the $MEAN = 1.2$ of the group. The variable was normally distributed ($p > 0.05$) for each group, as assessed by Shapiro–Wilk's test of normality. We can assume the homogeneity of variances in the different proactive conditions ($p > 0.05$ by Levene's test).

The results of the one-way ANOVA test on perceived creativity of the robot and our 3 conditions are given as (*ns*), $F(2, 27) = 2.28$, $p = 0.12$, $ges = 0.14$, as presented in **Figure 5** (where F is the result of the test, and ges is the generalized effect size). Given the value of p, we cannot conclude on the difference between the group conditions and the perceived creativity of the robot. As shown in the figure, the mean and standard deviation (SD) of the conditions are the following: high proactive condition with $MEAN = 1.3$, SD $= 0.90$, medium proactive condition with $MEAN = 1.2$, SD $= 1.08$, and no proactive condition with $MEAN = 0.3$, SD $= 1.27$. What is interesting to see is that the means of all three conditions of the robot's proactive behavior are perceived as creative (with values above 0 in the UEQ). The two levels of proactivity (medium–high) are perceived as more creative ($MEAN = 1.2$ and $MEAN = 1.3$, respectively) than the no proactive condition ($MEAN = 0.3$). In the case of the no proactive condition, the mean is close to zero with $MEAN = 0.3$, SD $= 1.27$, suggesting that participants may not have been able to assign a clear verdict about creativity in the robot's behavior. This shows that there is some perceivable difference to the user between the no proactive conditions and the proactive conditions, though not statistically significant according to the method used. Looking deeper into the generalized effect size, we see that $ges = 0.14$ (14%). This means that 14% of the change in the perceived creativity of the robot could be affected by the proactive conditions.

Thus, even without a statistically significant difference (according to this method), as part of an exploratory analysis, we looked at the means of each condition and effect size and

[2]Based on the practices as outlined in this resource https://www.datanovia.com/en/lessons/anova-in-r/#check-assumptions.

found that there still could be links with the proactivity of the robot and the perceived creativity of the robot. It is plausible that the users hesitated to rate the robot's creativity as dull so as to not seem harsh, given the positively skewed labels. During the post-experiment interviews, the participants indicated that the simple acknowledgment from the robot (the knowledge state marker "oh") was seen as better than no acknowledgment at all. For future work, it may be better to consider no verbal feedback whatsoever from the robot to test no proactivity.

Additionally, we did not find a statistically significant difference between the two levels of proactivity according to this method (medium with MEAN = 1.2, SD = 1.08 and full with MEAN = 1.3, SD = 0.90). Therefore, it is too early to establish any relationship between the frequency of proactive behavior and the scale of perceived creativity in the behavior. Therefore, H1 is not completely supported, in the sense that both parts are not validated (H1: proactivity and perceived creativity of the robot; the proactivity of the robot behavior will affect the perceived creativity. Proactivity in the robot behavior and its perceived creativity are related). The exploratory results suggest that the proactive condition is affecting the perceived creativity of the robot. But we could find statically significant links between the perceived creativity of the robot and the proactive condition robots to establish any relation. That is why there is a need for further investigation in this direction.

5.2 Observed Creativity in the User

To further explore the factors associated with the user's creativity, we conducted various analyses on the quantitative data from the study, such as the following: how many recipes were completed successfully? How many new recipes were created?

The design of the experiment was to encourage participants to complete at least one dish in phase 1 and then to complete or create at least one dish in phase 2. After that, participants were free to continue further iterations of phase 2 and complete or create more dishes. Participants can also skip a phase without completing or creating a dish.

We conduct a one-way ANOVA test to see if the total number of dishes that the user completed was different in our 3 group conditions. We run tests before conducting the one-way ANOVA test on the dependent variable (total number of completed dishes) to check that the assumptions are met. We have three outliers which are extreme; three users under the full condition completed (2,5,2) dishes compared to the *MEAN* = 3.0 of the group. The variable was normally distributed ($p > 0.05$) for medium and no proactive group conditions but was not normally distributed ($p < 0.05$) for the full proactive group condition, as assessed by Shapiro–Wilk's test of normality. We can assume the homogeneity of variances in the different proactive conditions ($p > 0.05$ by Levene's test). The results of the one-way ANOVA test to see if the total number of dishes that the user completed was different in our 3 group conditions are given as (*ns*), $F(2, 27) = 0.10$, $p = 0.9$, $ges = 0.007$ as presented in **Figure 6**. Given the value of p, we cannot conclude on the difference between the group conditions and the total number of dishes that the user completed. As shown in the figure, an almost straight trend line is observed between the conditions of the mean and standard deviation (SD) as follows: high proactive condition with *MEAN* = 3.0, SD = 0.81, medium proactive condition with *MEAN* = 3.1, SD = 1.28, and no proactive condition with *MEAN* = 3.2, SD = 0.78.

However, we found interesting observations when we conducted a one-way ANOVA test to see if the number of new dishes created was different in our 3 group conditions. We run tests before conducting the one-way ANOVA test on the dependent variable (number of created new dishes) to check that the assumptions are met. We have two outliers which are extreme; two users under the full condition created (1,2) new dishes compared to the *MEAN* = 0.3 of the group. The variable was normally distributed ($p > 0.05$) for medium and no proactive group conditions but was not normally distributed ($p < 0.05$) for the full proactive group condition, as assessed by Shapiro–Wilk's test of normality. We can assume the homogeneity of variances in the different proactive conditions ($p > 0.05$ by Levene's test). The results of the one-way ANOVA test to see if the number of new dishes created was different in our 3 group conditions are given as (***), $F(2, 27) = 10.62$, $p < 0.0003$, $ges = 0.44$ as shown in **Figure 7**. This graph is related to phase 2 of the experiment, where the participants were free to converge toward a dish from the list or proceed toward creating a new dish. Given the value of p, creating new dishes had a statistically significant difference between the proactive behavior conditions. We followed up with *post hoc* tests (t-test) to multiple pairwise comparisons between groups. It can be seen from **Figure 7** that there is a statistically significant difference between the no proactive condition and the full proactive condition with $p = 0.000019$(***) and between the no proactive and the medium proactive condition with $p = 0.0045$(**). There is no statistically significant difference (according to this method) between no and medium proactive conditions with $p = 0.7$(*ns*). Thus, it is observed that the number of new dishes which is created per person is significantly lower in full proactive conditions than in no and medium proactive conditions. Even in the no proactive condition with *MEAN* = 2.2, SD = 0.78 and the medium proactive condition with *MEAN* = 2, SD = 1.41, it shows that the number of new dishes which is created per person is lower in the medium proactive condition than in the no proactive condition. Hence, the analysis results support our hypothesis H2 (H2: proactivity and the user's creativity). Once the robot is very proactive and heavily interrupting the user toward achieving a goal, participants can complete the task (as shown in **Figure 6**) but are not flexible and free enough to create new recipes, as shown in **Figure 7**, hence being less creative.

Another interesting observation is that the medium proactivity condition has the maximum number of created new dishes per person (at most 5), whereas for no proactivity, most of the participants stopped after creating a maximum of 3 new dishes. In medium proactivity conditions, the highest number of new dishes created per person is observed. This is fascinating and suggests a need for a balance. It hints that balanced proactivity could encourage prolonged creativity. It needs further studies to define the boundaries of the balanced proactivity.

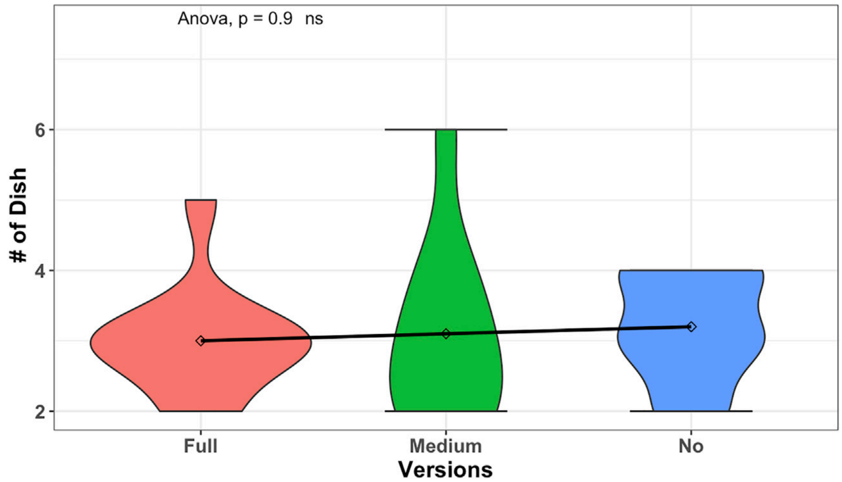


FIGURE 6 | Comparison results of total number of completed or created recipes to the effects of different versions of robot behavior to be creative on completing or creating recipes by increasing number of recipes that is created in total is not shown as significantly changed.

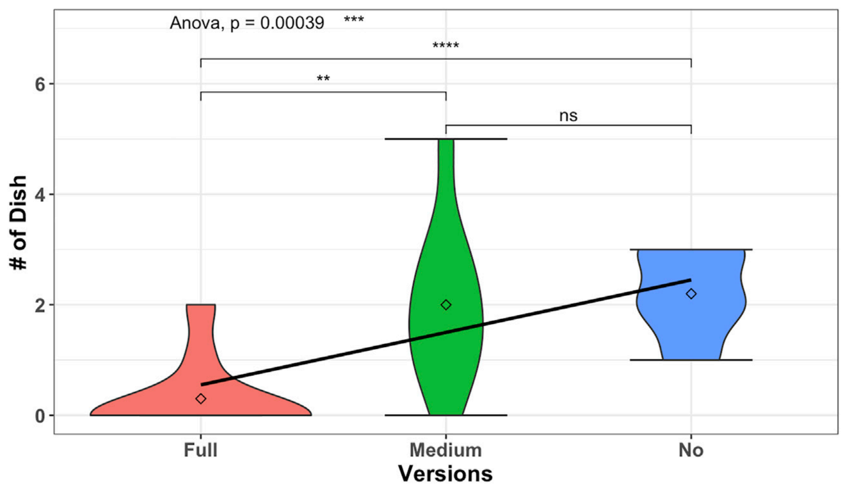


FIGURE 7 | Comparison results of creating new recipe to the effects of different versions of robot behavior to be creative on creating new recipes. It clearly shows that the balanced proactivity (version of medium proactive robot) is supporting a greater number of new recipes to be created by the users.

In summary, there is no strong observation about different frequencies of proactive behavior on constructing a recipe. From **Figure 6**, it is shown from the bump in full and no proactive conditions that the majority of the people tend to create three dishes. So, the proactivity has not affected their motivation of creating recipes. However, it is also observed that when there is a space between interruptions, it kind of encourages the users to play more. On the other hand, when the robot proactively creates suggestions for the users, the users' creativity decreases. The users tend to follow the robot suggestions and reduce their creativity process. As shown in the no proactivity case, since there is no help from the robot, users tend to be creative on constructing a recipe. However, in medium and, even heavily, in full proactive cases, when the robot is starting to help, the user's flexibility for being creative seems to be reducing, as the users mostly go with the flow that the robot suggested. However, further study is needed to explain the reason for the changes in the user's behavior.

5.3 Goal Achievement and Proactivity

To explore the benefits of proactive behavior on task accomplishment, we focus on phase 1 of the task and conducted the analysis of the comparison between each condition. How many times have recipes been done successfully? (see **Figure 8**) and how much time does the user spend while reaching the successful result? (see **Figure 9**) are used for analysis.

Figure 8 shows the successful completion of phase 1 of the experiment by the participants. As we can recall, in phase 1, the

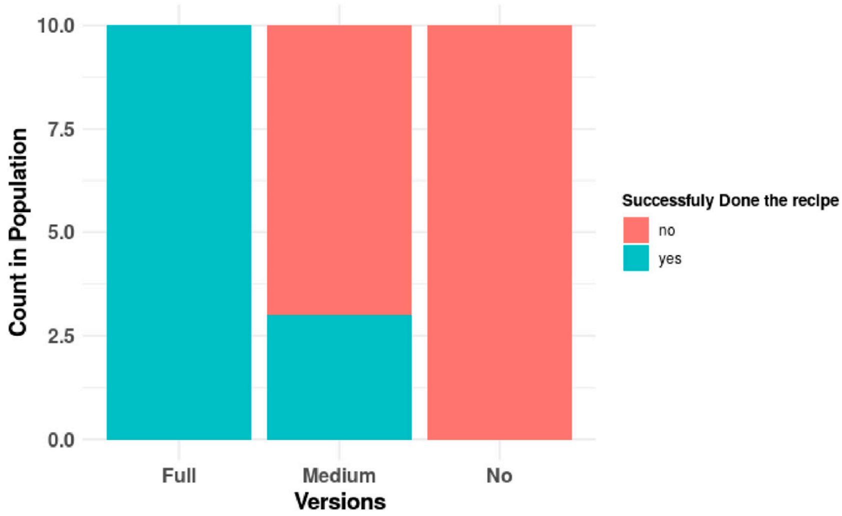


FIGURE 8 | Distribution of successfully achieving to assigned dish; graph groups the counts of number of participants who achieved the assigned goal successfully during phase 1 to the experiment. It shows the absolute dominance of success in full proactive behavior and failure in no proactive behavior of the robot. In the medium proactive behavior of the robot, variations were observed to have success or failure.

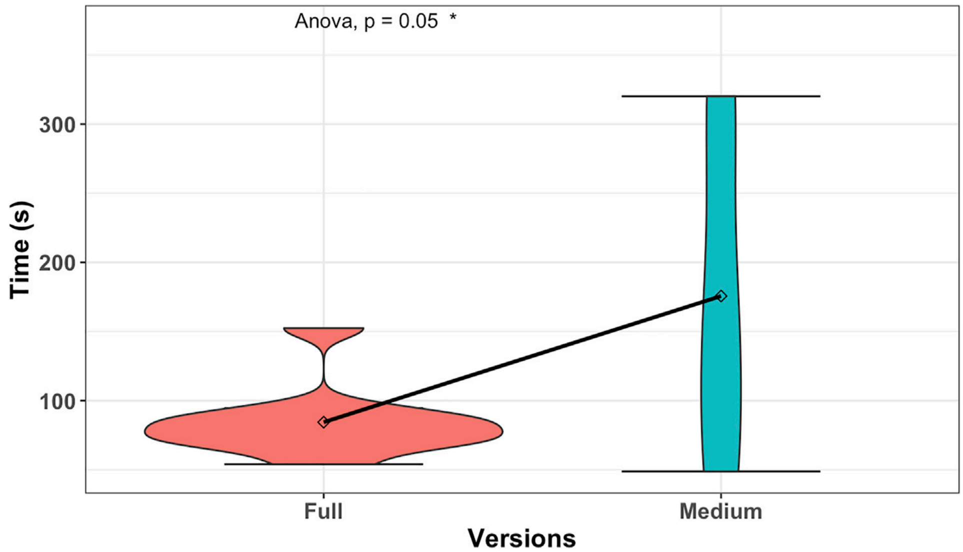


FIGURE 9 | Detailed analysis of time spent on the correct result of different levels of proactive behavior; the graph aims to show the difference between how much time the users spend while reaching the correct results in phase 1.

goal is assigned, and the target goal is known to the robot and the user. Hence, there is a joint goal to achieve. As we can see, all the participants exposed to the full proactive robot have successfully completed phase 1, whereas all the participants of the no proactive robot condition have failed. Furthermore, the failure rate was less in the medium proactivity condition, which shows 30% of success and 70% of failure. Pearson's chi square test of independence is applied to statistically analyze the correlation between different proactive behavior conditions and successfully completing phase 1. The result shows that there is significant relation between different conditions and success, $X^2(2, N = 30) = 21.44$, $p < 0.000022$(***). This supports our hypothesis H3 (H3: proactivity and goal achievement; there is a relationship between the robot's proactivity and the success of the HRI task. That is, the proactive behavior of the robot can help to achieve the goal of the task).

Furthermore, we conduct a one-way ANOVA test to see if the time spent between participants who achieved phase 1 successfully was different in group conditions. We run tests before conducting the one-way ANOVA test on the dependent

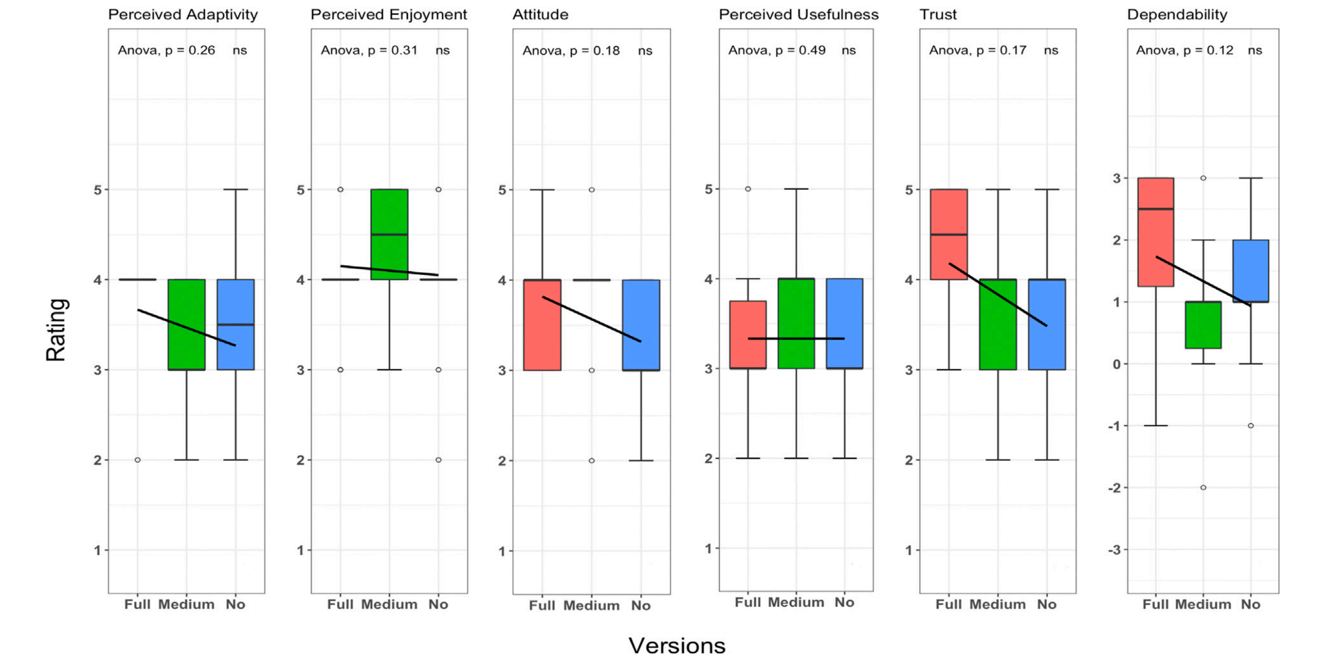


FIGURE 10 | Analysis of questionnaire; the graph visualizes the united results of the questionnaire with ANOVA and *post hoc t*-test analysis. The scale is a 5-point Likert scale (ALMERE), except for Dependability, which was part of another subset of the questionnaire using a 7-point Likert scale (UEQ).

variable (time spent) to check that the assumptions are met. We only have one extreme outlier; one user in the full proactive condition spent 152 s to reach success compared to the *MEAN* = 84.38 of the group. The variable was normally distributed ($p > 0.05$) for each (full and medium) condition. We can assume the homogeneity of the variances in different proactive conditions ($p > 0.05$ by Levene's test). The results of the one-way ANOVA test on time spent between participants who achieved phase 1 successfully and our proactive conditions are given as (*), $F(1, 11) = 4.84$, $p = 0.05$, $ges = 0.3$ as shown in **Figure 9**. Given the value of p, we can observe a significant difference in time spent between participants who achieved phase 1 successfully. It should be noted that phase 1 does not include the creation of new recipes. Therefore, these two findings combined also indicate that participants are more successful and spend less time in reaching the goal, with robots having a higher frequency of proactive behaviors.

5.4 Proactivity Level and Effects on Perceived Attributes

Figure 10 shows the overall impression of the participants about the robot's behavior in different versions. Although we did not find statistically significant differences (according to this method) to reach any conclusion or establish any solid relation for each scale, we are pointing out some of the findings for further investigation. The summary of the analysis for each scale is as follows:

Perceived adaptivity is one of the scales of the ALMERE questionnaire that measures users' perception of providing appropriate support by the robot. The user is asked to rate from 1: Do not Agree to 5: Totally Agree, with the statement "*I think the robot will help me when I consider it to be necessary*" on a 5-point Likert scale. We conduct a one-way ANOVA test to see if the perceived adaptivity of the robot was different in our 3 group conditions. We run tests before conducting the one-way ANOVA test on the dependent variable (perceived adaptivity) to check that the assumptions are met. We only have one extreme outlier; one user in the medium condition rated the perceived adaptivity of the robot as less not agreed (2) compared to the *MEAN* = 3.2 of the group. The variable was normally distributed ($p > 0.05$) for medium and no proactive conditions but was not normally distributed ($p < 0.05$) for the full proactive condition, as assessed by Shapiro–Wilk's test of normality. We cannot assume the homogeneity of variances in the different proactive conditions ($p < 0.05$ by Levene's test). The results of the one-way ANOVA test on perceived adaptivity of the robot and our 3 conditions are given as (*ns*), $F(2, 27) = 1.43$, $p = 0.25$, $ges = 0.09$. Given the value of p, we cannot conclude on the difference between the group conditions and the perceived adaptivity of the robot. As shown in the figure, the mean and standard deviation (SD) of the conditions are the following: high proactive condition with *MEAN* = 3.8, SD = 0.63, medium proactive condition with *MEAN* = 3.2, SD = 0.78, and no proactive condition with *MEAN* = 3.4, SD = 0.96. In that sense, it is observed that participants found the full proactive condition of the robot the most adaptable. However, it is interesting to see the robot which did not give any suggestions be seen as more adaptable than the robot which is giving sparse suggestions (in the medium proactive condition). It might be because of various factors ranging from frustration of not getting enough suggestions (in case of the medium proactive condition) to the robot acknowledging behaviors being seen as completely supporting to the user action (in the no proactive condition). Therefore, this is another interesting direction for further investigations.

Perceived enjoyment is one of the scales of the ALMERE questionnaire that measures the level of enjoyment of the user while interacting with the robot. The user is asked to rate from 1: Do not Agree to 5: Totally Agree, with the statement "*I enjoyed the robot talking to me*" on a 5-point Likert scale. We conduct a one-way ANOVA test to see if the perceived enjoyment of the robot was different in our 3 group conditions. We run tests before conducting the one-way ANOVA test on the dependent variable (perceived enjoyment) to check that the assumptions are met. We have eight extreme outliers; four users in the full proactive condition, two of whom rated the perceived enjoyment of the robot as totally agree (5) and the other two rated as less not agree (2) compared to the $MEAN = 4.00$ of the group, and four users in the no proactive condition, two of whom rated the perceived enjoyment of the robot as totally agree (5), one of whom rated slightly agree (3), and the other rated less not agree (2) compared to the $MEAN =$ of the group. The variable was normally distributed ($p > 0.05$) for full and no proactive conditions but was not normally distributed ($p < 0.05$) for the medium proactive condition, as assessed by Shapiro–Wilk's test of normality. We can assume the homogeneity of variances in the different proactive conditions ($p > 0.05$ by Levene's test). The results of the one-way ANOVA test on perceived enjoyment of the robot and our 3 conditions are given as (ns), $F(2, 27) = 1.23$, $p = 0.30$, $ges = 0.08$. Given the value of p, we cannot conclude on the difference between the group conditions and the perceived enjoyment of the robot. As shown in the figure, the mean and standard deviation (SD) of the conditions are the following: high proactive condition with $MEAN = 4.00$, SD $= 0.66$, medium proactive condition with $MEAN = 4.4$, SD $= 0.69$, and no proactive condition with $MEAN = 3.9$, SD $= 0.87$. The results show that participants found the medium proactive behavior condition of the robot to be a more enjoyable companion. This might be because such behavior might not constrain the flow of the task very much with overload of suggestions or not seem engaged enough because of no suggestion.

Attitude is one of the scales of the ALMERE questionnaire that measures the user's attitude toward the particular technology behind the version of robot behavior they have been exposed to. The user is asked to rate from 1: Do not Agree to 5: Totally Agree, with the statement "*The robot would make my life more interesting*" on a 5-point Likert scale. We conduct a one-way ANOVA test to see if the attitude toward the robot was different in our 3 group conditions. We run tests before conducting the one-way ANOVA test on the dependent variable (attitude) to check that the assumptions are met. We only have three extreme outliers; three users in the medium condition, one of whom rated the attitude toward the robot as less not agree (2), one of whom rated as slightly agree (3), and the other rated as totally agree (5) compared to the $MEAN = 3.8$ of the group. The variable was normally distributed ($p > 0.05$) for full and no proactive conditions but was not normally distributed ($p < 0.05$) for the medium proactive condition, as assessed by Shapiro–Wilk's test of normality. We can assume the homogeneity of variances in the different proactive conditions ($p > 0.05$ by Levene's test). The results of the one-way ANOVA test on attitude toward the robot and our 3 conditions are given as (ns), $F(2, 27) = 1.82$, $p = 0.18$, $ges = 0.11$. Given the value of p, we cannot conclude on the difference between the group conditions and the attitude toward the robot. As shown in the figure, the mean and standard deviation (SD) of the conditions are the following: high proactive condition with $MEAN = 3.7$, SD $= 0.67$, medium proactive condition with $MEAN = 3.8$, SD $= 0.78$, and no proactive condition with $MEAN = 3.2$, SD $= 0.78$. The responses show that the medium proactive robot behavior is the most appreciated behavior.

Perceived usefulness is one of the scales of the ALMERE questionnaire that is another key aspect about the relevance of a particular behavior of the robot. The user is asked to rate from 1: Do not Agree to 5: Totally Agree, with the statement "*I think the robot can help me with many things*" on a 5-point Likert scale. We conduct a one-way ANOVA test to see if the perceived usefulness of the robot was different in our 3 group conditions. We run tests before conducting the one-way ANOVA test on the dependent variable (perceived usefulness) to check that the assumptions are met. We only have one outlier which is not extreme; one user in the full condition rated the perceived usefulness of the robot as totally agree (5) compared to the $MEAN = 3.2$ of the group. The variable was normally distributed ($p > 0.05$) for each condition, as assessed by Shapiro–Wilk's test of normality. We can assume the homogeneity of variances in the different proactive conditions ($p > 0.05$ by Levene's test). The results of the one-way ANOVA test on perceived usefulness of the robot and our 3 conditions are given as (ns), $F(2, 27) = 0.73$, $p = 0.48$, $ges = 0.05$. Given the value of p, we cannot conclude on the difference between the group conditions and the perceived usefulness of the robot. As shown in the figure, the mean and standard deviation (SD) of the conditions are the following: high proactive condition with $MEAN = 3.2$, SD $= 0.91$, medium proactive condition with $MEAN = 3.6$, SD $= 0.84$, and no proactive condition with $MEAN = 3.2$, SD $= 0.78$. Again, the responses show that the medium proactive robot behavior is preferred by the users.

Trust is one of the scales of the ALMERE questionnaire that measures the user intentions to comply with the robot's advice. The user is asked to rate from 1: Do not Agree to 5: Totally Agree, with the statement "*I would follow the advice the robot gives me*" on a 5-point Likert scale. We conduct a one-way ANOVA test to see if the trust toward the robot was different in our 3 group conditions. We run tests before conducting the one-way ANOVA test on the dependent variable (trust) to check that the assumptions are met. We do not have an outlier. The variable was normally distributed ($p > 0.05$) for medium and no proactive conditions but was not normally distributed ($p < 0.05$) for the full proactive condition, as assessed by Shapiro–Wilk's test of normality. We can assume the homogeneity of variances in the different proactive conditions ($p > 0.05$ by Levene's test). The results of the one-way ANOVA test on trust toward the robot and our 3 conditions are given as (ns), $F(2, 27) = 1.92$, $p = 0.16$, $ges = 0.12$. Given the value of p, we cannot conclude on the difference between the group conditions and the trust toward the robot. As shown in the figure, the mean and standard deviation (SD) of the conditions are the following: high proactive condition with $MEAN = 4.3$, $SD = 0.82$, medium proactive condition with $MEAN = 3.6$, $SD = 1.07$, and no proactive condition with $MEAN = 3.6$, $SD = 0.84$. The results show that even the full proactive condition lead the user to rely more on the robot with the full proactive condition.

Dependability is one of the scales of UEQ (User Experience Questionnaire) that measures how much the reactions to the robot's behavior are predictable. The user is asked to rate from dull to dependable, with the statement "*In my opinion, the reactions of the robot's behavior to my input and command is –*" on a 7-point scale (-3:Unpredictable to 3:Predictable). We conduct a one-way ANOVA test to see if the dependability of the robot was different in our 3 group conditions. We run tests before conducting the one-way ANOVA test on the dependent variable (dependability) to check that the assumptions are met. We have three outliers which are not extreme; two users in the medium proactive condition rated the dependability of the robot one as (-2) and one as dependable (3) compared to the $MEAN = 0.80$ of the group and one user in the no proactive condition rated the dependability of the robot as (-1) compared to the $MEAN = 1.20$ of the group. The variable was normally distributed ($p > 0.05$) for medium and no proactive conditions but was not normally distributed ($p < 0.05$) for the full proactive condition, as assessed by Shapiro–Wilk's test of normality. We can assume the homogeneity of variances in the different proactive conditions ($p > 0.05$ by Levene's test). The results of the one-way ANOVA test on dependability of the robot and our 3 conditions are given as (*ns*), $F(2, 27) = 2.3$, $p = 0.11$, $ges = 0.14$, (where F is the result of the test, and *ges* is the generalized effect size). Given the value of p, we cannot conclude on the difference between the group conditions and the trust toward the robot. As shown in the figure, the mean and standard deviation (SD) of the conditions are the following: high proactive condition with $MEAN = 2.0$, $SD = 1.33$, medium proactive condition with $MEAN = 0.8$, $SD = 1.31$, and no proactive condition with $MEAN = 1.2$, $SD = 1.13$. The initial findings suggest that the participants listened more to the robot, which generated more advice than the full proactive condition of the robot.

It will be interesting to investigate further in these directions to find the factors behind these observations and to further explore the right level of proactivity for the interaction to be more enjoyable, adaptive, useful, and establishing the necessary trust and dependability at the same time.

Such differences in the perception of different attributes in different versions of robot behavior support our hypothesis H4: proactivity level and user perception different levels of proactivity of the robot will have different user experiences on the perceived attributes.

6 DISCUSSION ON QUALITATIVE OBSERVATIONS

The interaction with the robot was not always so smooth. There were some problems related to the robot's vocal feedback such as some participants confusing the verb "egg" with "ice." So, they spent more time on understanding the robot's suggestion.

There were some cases in which the dish's name was the same as that known to the robot, but the participants selected different ingredients to create their own version of the dish. Those cases need to be investigated in future studies. However, it created an interesting interaction pattern as follows:

Robot: I thought you are selecting ingredients for < *..dish..* > but I don't know this recipe.
Robot: Could you please tell me the name of it?
Participant: I know but this is my < *..dish..* > that's why it's different.

In the current analysis, if the dish's ingredients are different, it is classified as a new dish since creativity assessment depends on knowledge. We classify the novelty of recipe creation according to what is provided by the task and what is known from the robot. It is important not to forget that the robot could only help with the limitation of its knowledge.

Participants of experiments are the employees of SoftBank Robotics Europe. They had experience with the Pepper robot. However, their background is mixed between technical (hardware and software) and non-technical (marketing, communication, and welcome desk). Nevertheless, this can introduce bias in terms of a more positive attitude toward the robot. In the future, we aim to experiment with more diverse users, hopefully, once the Covid-19 restrictions are over.

Some participants listened to the robot's feedback for the first phase but not very much during the second phase onward. Later, they stated, "*I knew what I was doing, so I did not listen to the robot's advice*" or "*I already asked the robot for help, it did not help me. Then, it offered some help. This time, I refused it.*" Such feedback indicates that in addition to considering the goal and future needs, the robot should also incorporate social signals and some aspects of reactiveness while generating its proactive behavior. It will be interesting to explore such factors and develop an inclusive computational model for behavior generation.

The effect of agency and embodied presence of the robot was observed strongly. For example, some participants perceived the "oh" response as positive, while others perceived negatively as respond proved as a neutral response to not reflect any opinion. It is expected from the extraction of previous research (Heritage, 1984) that involuntary interruption means anything. Some participants also think that the robot is enjoying the selections, so they continue to create a recipe process. Participants were so eager to get any reaction from the robot that they tried to put different naming. Some participants also played tricks to validate their perception about "oh" behavior at a medium proactive condition of the robot. This suggests that even some involuntary interruptions will keep participants motivated in a task, which might contribute to their prolonged creative "experiments." We believe that these differences in perception are related to participants' tendency to extract the meaning of each noise from the robot. It is not incontrovertible of the familiarity of participants with the robot.

This is another exciting direction for further studying the connection between robots' behavior and its effect on creativity in the user.

It is observed that sometimes the ingredients were limited to create an entirely new recipe. In those cases, the task-oriented proactive verbal communicative actions of the robot also confused some participants, as they stated that "I was not sure should I create a new recipe or try to create one of the given ones." Also, as some of the participants mentioned that they were not very good with cooking and recipe knowledge, that might be contributing to participants following the feedback from the robot. Such observations need further investigation on understanding the more in-depth relation between proactivity and creativity in an open-ended and domain-independent scenario.

7 CONCLUSION

This study attempts to explore the behavioral aspect of creativity in robots in the context of human–robot interaction. We hypothesized the dimension of bringing novelty in behavior by proactive actions by letting the robot initiate a suggestive interaction for a task that humans are supposed to do. We have presented the creative cooking experiment and analysis with the humanoid robot, Pepper. As this is an exploratory study, the preliminary finding hints toward the proactive behaviors of the robot somewhat affecting the perceived creativity of the robot. We have also provided pointers such as proactive behaviors not only leading users but also helping to keep achieving the goal of the task. We have shown that different levels of proactive behaviors have different effects and relations with various aspects of perceived attributes. To our knowledge, this is the first study of its kind on understanding the creativity and proactivity aspects together in a human–robot interaction context, from the perspective of achieving a goal and from the perspective of supporting creativity in the user. We have discussed and pointed out various aspects needing further investigation to strengthen our knowledge in this domain, including the finding that there seem to be trade-offs to find the right level of proactivity that will help to achieve the goal but leave space for the user to be creative, which we think is very important for the real-world deployment of social robots in day-to-day tasks and companionship.

7.1 Limitation of the Study

Due to the COVID-19 lockdown in France, it was not possible to conduct a physical experiment with potential end users. That is why we conduct a physical experiment with SoftBank Robotics Europe employees who have special allowance to enter the working area.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SB designed the architecture, implemented the code, ran the experiments, and wrote the manuscript draft. MAC facilitated the experiment inside the company and, with AP, supervised the research, improved the manuscript, and provided insights and guidance, and MOC supervised the research and provided insights and guidance.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 765955 (ANIMATAS).

ACKNOWLEDGMENTS

We wanted to thank the SoftBank Robotics Europe employees who participated in the study.

REFERENCES

Ali, S., Moroso, T., and Breazeal, C. (2019). "Can Children Learn Creativity from A Social Robot?," in C&C '19: Proceedings of the 2019 on Creativity and Cognition, San Diego, CA, June 23–26, 2019 (New York, NY: ACM–Association for Computing Machinery), 359–368. doi:10.1145/3325480.3325499

Alves-Oliveira, P., Tulli, S., Wilken, P., Merhej, R., Gandum, J., and Paiva, A. (2019). Sparking Creativity with Robots: A Design Perspective," in Robots for Social Good: Exploring Critical Design for HRI Workshop at 14th ACM/IEEE International Conference on Human-Robot Interaction, HRI (2019), Daegu, South Korea. doi:10.31219/osf.io/za5h8

Apiola, M., Lattu, M., and Pasanen, T. A. (2010). "Creativity and Intrinsic Motivation in Computer Science Education," in ITiCSE '10: Proceedings of the Fifteenth Annual Conference on Innovation and Technology in Computer Science Education (New York, NY, USA: Association for Computing Machinery), 199–203. doi:10.1145/1822090.1822147

Awais, M., and Henrich, D. (2012). "Proactive Premature Intention Estimation for Intuitive Human-Robot Collaboration," in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 7-12 Oct. 2012 (Vilamoura-Algarve, Portugal: IEEE). doi:10.1109/IROS.2012.6385880

Baas, M., De Dreu, C. K. W., and Nijstad, B. A. (2008). A Meta-Analysis of 25 Years of Mood-Creativity Research: Hedonic Tone, Activation, or Regulatory Focus? *Psychol. Bull.* 134, 779–806. doi:10.1037/a0012815

Bader, S., Nicolay, R., and Kirste, T. (2013). "Agent-based Proactive Support in Smart Environments," in Proceedings - 9th International Conference on Intelligent Environments, IE 2013, Athens, Greece, July 16–17, 2013 (IEEE), 220–223. doi:10.1109/IE.2013.30

Baraglia, J., Cakmak, M., Nagai, Y., Rao, R., and Asada, M. (2016). "Efficient Human-robot Collaboration: When Should a Robot Take Initiative ?," in The International Journal of Robotics Research, February 5, 2017 (SAGE Journals), 563–579. doi:10.1177/0278364916688253

Bateman, T. S., and Crant, J. M. (1999). Proactive Behavior: Meaning, Impact, Recommendations. *Business Horizons* 42, 63–70. doi:10.1016/s0007-6813(99)80023-8

Bateman, T. S., and Crant, J. M. (1993). The Proactive Component of Organizational Behavior: A Measure and Correlates. *J. Organiz. Behav.* 14, 103–118. doi:10.1002/job.4030140202

Bremner, P., Dennis, L. A., Fisher, M., and Winfield, A. F. (2019). On Proactive, Transparent, and Verifiable Ethical Reasoning for Robots. *Proc. IEEE* 107, 541–561. doi:10.1109/JPROC.2019.2898267

Bussy, A., Gergondet, P., Kheddar, A., Keith, F., and Crosnier, A. (2012). "Proactive Behavior of a Humanoid Robot in a Haptic Transportation Task with a Human Partner," in Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, Paris, France, September 9–13, 2012 (IEEE), 962–967. doi:10.1109/roman.2012.6343874

Buyukgoz, S., Chamoux, M., Pandey, A. K., and Chetouani, M. (2020). "Exploring Proactivity Aspect of Creativity in Human-Robot Interaction*," in The Twelfth International Conference of Social Robotics (ICSR), Workshop of Creativity and Robotics (New York, NY, USA: Apress).

Cramer, H. S. M., Kemper, N. A., Amin, A., and Evers, V. (2009). "The Effects of Robot Touch and Proactive Behaviour on Perceptions of Human-Robot Interactions," in Proceedings of the 4th ACM/IEEE international conference on Human robot interaction - HRI '09 (New York, New York, USA: ACM Press), 275. doi:10.1145/1514095.1514173

Csikszentmihalyi, M. (2009). *Creativity: Flow and the Psychology of Discovery and Invention.*HarperCollins e-books. October 13, 2009.

De Dreu, C. K. W., Baas, M., and Nijstad, B. A. (2008). Hedonic Tone and Activation Level in the Mood-Creativity Link: Toward a Dual Pathway to Creativity Model. *J. Personal. Soc. Psychol.* 94, 739–756. doi:10.1037/0022-3514.94.5.739

Elgarf, M., Calvo-Barajas, N., Paiva, A., Castellano, G., and Peters, C. (2021). "Reward Seeking or Loss Aversion?: Impact of Regulatory Focus Theory on Emotional Induction in Children and Their Behavior Towards a Social Robot Published," in CHI 21: Proceedings of the 2021 CHI Conference on Human Factors in Computing System, Yokohama, Japan, May 8–13, 2021 (ACM - Association for Computing Machinery), 1–11. doi:10.1145/3411764.3445486

Fiore, M., Khambhaita, H., Milliez, G., Alami, R., and Adaptive, R. A. A. (2015). "An Adaptive and Proactive Human-Aware Robot Guide," in International Conference on Social Robotics, Paris, France, October 26–30, 2015 (Springer Cham) 9388, 194–203. doi:10.1007/978-3-319-25554-5_20

Franken, R. (1994). *Human Motivation*. Pacific Grove, CA: Brooks/Cole Publishing Company.

Garrell, A., Villamizar, M., Moreno-Noguer, F., and Sanfeliu, A. (2013). "Proactive Behavior of an Autonomous mobile Robot for Human-Assisted Learning," in Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, Gyeongju, Korea (South), August 26–29, 2013 (IEEE), 107–113. doi:10.1109/ROMAN.2013.6628463

George, J. M., and Zhou, J. (2007). Dual Tuning in a Supportive Context: Joint Contributions of Positive Mood, Negative Mood, and Supervisory Behaviors to Employee Creativity. *Acad. Manage. J.* 50, 605–622. doi:10.5465/amj.2007.25525934

Goncalo, J. (2019). Matthew a. Cronin and Jeffrey Loewenstein: The Craft of Creativity. *Administrative Sci. Q.* 64, NP38–NP40. doi:10.1177/0001839219863923

Gong, Z., and Zhang, N. (2017). Using a Feedback Environment to Improve Creative Performance: A Dynamic Affect Perspective. *Front. Psychol.* 8, 1398. doi:10.3389/fpsyg.2017.01398

Grant, A. M., and Ashford, S. J. (2008). The Dynamics of Proactivity at Work. *Res. Organizational Behav.* 28, 3–34. doi:10.1016/j.riob.2008.04.002

Grosinger, J., Pecora, F., and Saffiotti, A. (2016). "Making Robots Proactive through Equilibrium Maintenance," in IJCAI International Joint Conference on Artificial Intelligence, New York, NY, July 9–15, 2016 (AAAI Press), 3375–3381.

Han, Z., and Yanco, H. (2019). "The Effects of Proactive Release Behaviors during Human-Robot Handovers," in ACM/IEEE International Conference on Human-Robot Interaction, Daegu, Korea (South), March 11–14, 2019 (IEEE), 440–448. doi:10.1109/hri.2019.8673085

Heerink, M., Kröse, B., Evers, V., and Wielinga, B. (2010). Assessing Acceptance of Assistive Social Agent Technology by Older Adults: The Almere Model. *Int. J. Soc. Robotics* 2, 361–375. doi:10.1007/s12369-010-0068-5

Heritage, J. (1984). "A Change-Of-State Token and Aspects of its Sequential Placement," in *Structures of Social Action: Studies in Conversation Analysis*. Editors J. M. Atkinson and J. Heritage (Cambridge, U.K.: Cambridge University Press), Chap. 13, 299–345.

Hu, Y., Feng, L., Mutlu, B., and Admoni, H. (2021). "Exploring the Role of Social Robot Behaviors in a Creative Activity," in DIS 21: Designing Interactive Systems Conference 2021, Virtual Event United States, June 28–July 2, 2021 (New York, NY: ACM - Association for Computing Machinery), 1380–1389. doi:10.1145/3461778.3462116

Joo, B.-K. B., and Bennett, R. H., III (2018). The Influence of Proactivity on Creative Behavior, Organizational Commitment, and Job Performance: Evidence from a Korean Multinational. *J. Int. Interdiscip. Business Res.* 5, 1–20.

Kahn, P. H., Kanda, T., Ishiguro, H., Gill, B. T., Shen, S., Ruckert, J. H., and Gary, H. E. (2016). "Human Creativity Can Be Facilitated through Interacting with a Social Robot," in ACM/IEEE International Conference on Human-Robot Interaction, Christchurch, NZ, March 7–10, 2016 (IEEE), 173–180. doi:10.1109/HRI.2016.7451749

Kraus, M., Wagner, N., and Minker, W. (2020). "Effects of Proactive Dialogue Strategies on Human-Computer Trust," in UMAP 2020 - Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (New York, NY, USA: Association for Computing Machinery, Inc), 107–116. doi:10.1145/3340631.3394840

Law, M. V., Jeong, J., Kwatra, A., Jung, M. F., and Hoffman, G. (2019). "Negotiating the Creative Space in Human-Robot Collaborative Design," in DIS '19: Proceedings of the 2019 on Designing Interactive Systems Conference, San Diego, CA, June 23–28, 2019 (New York, NY: ACM - Association for Computing Machinery), 645–657. doi:10.1145/3322276.3322343

Lemaignan, S., Ros, R., Mösenlechner, L., Alami, R., and Beetz, M. (2010). "ORO, a Knowledge Management Platform for Cognitive Architectures in Robotics," in IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings, Taipei, Taiwan, October 18–22, 2010 (IEEE), 3548–3553. doi:10.1109/iros.2010.5649547

Lin, Y., Guo, J., Chen, Y., Yao, C., and Ying, F. (2020). "It Is Your Turn: Collaborative Ideation with a Co-creative Robot through Sketch," in CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, April 25–30, 2020 (New York, NY: Association for Computing Machinery), 1–14. doi:10.1145/3313831.3376258

Liu, P., Glas, D. F., Kanda, T., and Ishiguro, H. (2018). Learning Proactive Behavior for Interactive Social Robots. *Auton. Robot* 42, 1067–1085. doi:10.1007/s10514-017-9671-8

Moroney, L., and Moroney, L. (2017). "The Firebase Realtime Database," in The Definitive Guide to Firebase (New York, NY, USA: Apress), 51–71. doi:10.1007/978-1-4842-2943-9_3

Moulin-Frier, C., Fischer, T., Petit, M., Pointeau, G., Puigbo, J. Y., Pattacini, U., et al. (2017). "DAC-h3: A Proactive Robot Cognitive Architecture to Acquire and Express Knowledge about the World and the Self," in IEEE Transactions on Cognitive and Developmental Systems, September 18, 2017 (IEEE), 1005–1022. doi:10.1016/j.aqrep.2017.11.005

Myers, K., and Yorke-Smith, N. (2007). "Proactive Behavior of a Personal Assistive Agent," in Proceedings of the AAMAS Workshop on Metareasoning in Agent-Based Systems, Honolulu, HI, 31–45.

Pandey, A. K., Ali, M., and Alami, R. (2013). Towards a Task-Aware Proactive Sociable Robot Based on Multi-State Perspective-Taking. *Int. J. Soc. Robotics* 5, 215–236. doi:10.1007/s12369-013-0181-3

Pandey, A. K., and Gelin, R. (2018). A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of its Kind. *IEEE Robot. Automat. Mag.* 25, 40–48. doi:10.1109/MRA.2018.2833157

Peng, Z., Kwon, Y., Lu, J., Wu, Z., and Ma, X. (2019). "Design and Evaluation of Service Robot's Proactivity in Decision-Making Support Process," in

Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19 (New York, New York, USA: ACM Press), 1–13. doi:10.1145/3290605.3300328

Ritter, S. M., Gu, X., Crijns, M., and Biekens, P. (2020). Fostering Students' Creative Thinking Skills by Means of a One-Year Creativity Training Program. *PLoS ONE* 15, e0229773. doi:10.1371/JOURNAL.PONE.0229773

Rivoire, C., and Lim, A. (2016). "Habit Detection within a Long-Term Interaction with a Social Robot," in DAA 2016 - Proceedings of the International Workshop on Social Learning and Multimodal Interaction for Designing Artificial Agents (New York, New York, USA: Association for Computing Machinery, Inc). doi:10.1145/3005338.3005342

Runco, M. A. (1992). Children's Divergent Thinking and Creative Ideation. *Develop. Rev.* 12, 233–264. doi:10.1016/0273-2297(92)90010-Y

Runco, M. A. (1993). Divergent Thinking, Creativity, and Giftedness. *Gifted Child. Q.* 37, 16–22. doi:10.1177/001698629303700103

Schrepp, M., and Thomaschewski, J. (2019). Design and Validation of a Framework for the Creation of User Experience Questionnaires. *Int. J. Interactive Multimedia Artif. Intelligence* 5, 88–95. doi:10.9781/ijimai.2019.06.006

Sternberg, R. J. (2003). Creative Thinking in the Classroom. *Scand. J. Educ. Res.* 47, 325–338. doi:10.1080/00313830308595

Thobbi, A., Gu, Y., and Sheng, W. (2011). "Using Human Motion Estimation for Human-Robot Cooperative Manipulation," in IEEE International Conference on Intelligent Robots and Systems, San Francisco, CA, September 25–30, 2011 (IEEE), 2873–2878. doi:10.1109/iros.2011.6094904

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and Sharing Intentions: The Origins of Cultural Cognition. *Behav. Brain Sci.* 28, 675–691. doi:10.1017/S0140525X05000129

Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking: Norms-Technical Manual*. Princeton, NJ: Personal Press, 1–79.

Ujjwal, K. C., and Chodorowski, J. (2019). A Case Study of Adding Proactivity in Indoor Social Robots Using Belief-Desire-Intention (BDI) Model. *Biomimetics (Basel)* 4, 74. doi:10.3390/biomimetics4040074

Wang, X., Ye, S., and Teo, H. (2014). "Effects of Interruptions on Creative Thinking," in Thirty Fifth International Conference on Information Systems, Auckland, December 15, 2014 (AIS Electronic Library (AISeL)).

Weisberg, R. (1993). *Creativity: Beyond the Myth of Genius. Books in Psychology*. New York: W. H. Freeman.

White, A., Tate, A., and Rovatsos, M. (2017). CAMP-BDI: A Pre-emptive Approach for Plan Execution Robustness in Multiagent Systems," in 18th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA 2015), Bertinoro, Italy, October 26–30, 2015 (Springer, Cham), 65–84.

Woodman, R. W., Sawyer, J. E., and Griffin, R. W. (1993). Toward a Theory of Organizational Creativity. *Acad. Manage. Rev.* 18, 293–321. doi:10.5465/amr.1993.3997517

Zhang, Y., Narayanan, V., Chakraborti, T., and Kambhampati, S. (2015). "A Human Factors Analysis of Proactive Support in Human-Robot Teaming," in IEEE International Conference on Intelligent Robots and Systems, Hamburg, Germany, September 28–October 2, 2015 (IEEE), 3586–3593. doi:10.1109/IROS.2015.7353878

Conflict of Interest: Author AP is engaged with the companies beingAI Limited and Socients AI and Robotics. Authors SB and MAC are employed by the company SoftBank Robotics Europe.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

TYPE Original Research
PUBLISHED 24 August 2022
DOI 10.3389/frobt.2022.695162

Check for updates

OPEN ACCESS

EDITED BY
Vasanth Sarathy,
Tufts University, United States

REVIEWED BY
Jill Dosso,
University of British Columbia, Canada
Patricia Alves-Oliveira,
University of Washington, United States

*CORRESPONDENCE
Eduardo Benítez Sandoval,
e.sandoval@unsw.edu.au

SPECIALTY SECTION
This article was submitted to Human-Robot Interaction,
a section of the journal
Frontiers in Robotics and AI

RECEIVED 14 April 2021
ACCEPTED 28 June 2022
PUBLISHED 24 August 2022

CITATION
Sandoval EB, Sosa R, Cappuccio M and Bednarz T (2022), Human–robot creative interactions: Exploring creativity in artificial agents using a storytelling game.
Front. Robot. AI 9:695162.
doi: 10.3389/frobt.2022.695162

Human–robot creative interactions: Exploring creativity in artificial agents using a storytelling game

Eduardo Benítez Sandoval[1]*, Ricardo Sosa[2,3], Massimiliano Cappuccio[4] and Tomasz Bednarz[5]

[1]Creative Robotics Lab, School of Arts and Design, Faculty of Arts, Design and Architecture, University of New South Wales, Sydney, Australia, [2]Faculty of Design ad Creative Technologies, Auckland University of Technology, Auckland, New Zealand, [3]Monash University, Design, Melbourne, Australia, [4]University of New South Wales, Canberra, Australia, [5]EPICentre and CSIRO's Data61, University of New South Wales, Sydney, Australia

Creativity in social robots requires further attention in the interdisciplinary field of human–robot interaction (HRI). This study investigates the hypothesized connection between the perceived creative agency and the animacy of social robots. The goal of this work is to assess the relevance of robot movements in the attribution of creativity to robots. The results of this work inform the design of future human–robot creative interactions (HRCI). The study uses a storytelling game based on visual imagery inspired by the game "Story Cubes" to explore the perceived creative agency of social robots. This game is used to tell a classic story for children with an alternative ending. A 2 × 2 experiment was designed to compare two conditions: the robot telling the original version of the story and the robot plot twisting the end of the story. A Robotis Mini humanoid robot was used for the experiment, and we adapted the Short Scale of Creative Self (SSCS) to measure perceived creative agency in robots. We also used the Godspeed scale to explore different attributes of social robots in this setting. We did not obtain significant main effects of the robot movements or the story in the participants' scores. However, we identified significant main effects of the robot movements in features of animacy, likeability, and perceived safety. This initial work encourages further studies experimenting with different robot embodiment and movements to evaluate the perceived creative agency in robots and inform the design of future robots that participate in creative interactions.

KEYWORDS

creative robots, robot games, story cubes, creative interactions, short-scale creative self scale, storytelling, the ugly duckling

1 Introduction

An important dimension in social interaction is how agents perceive the intelligence and creativity of other agents. However, creativity is an under-explored area in the study of human–robot interaction (HRI) (Saunders et al., 2013). Creativity can be defined as the capacity to imagine alternative futures. Despite its relevance in shaping everyday interactions, we have a limited knowledge of how creativity is perceived and attributed by humans to self and others. Furthermore, the display of creativity is a subjective phenomenon that is challenging to study using experimental methods of inquiry (Svanæs, 2013). This study is drawn from work in the area of self-assessment of creativity to evaluate the possible connections between the perceived creative agency and the animacy and kinesthetics of social robots.

In artificial agents such as social robots or screen-based avatars, the attribution of creativity has remained largely unaddressed in the field of HRI until recently (Alves-Oliveira et al., 2017, 2020). While artificial intelligence has been investigated extensively, and artificial creativity defined as the creativity attributed to artificial agents remains to be addressed particularly in experimental studies. Some creative behaviors have been simulated using language, pattern recognition, and evolutionary generative systems as shown (Pham et al., 2017; Gizzi et al., 2019; Myoo, 2019; Uneeq Ltd, Digital Humans, 2021; OpenAI, 2021). Similarly, some work has been performed on users' non-verbal behavior, self-presence, social presence, and interpersonal attraction in one-to-one human–agent interaction on collaborative virtual environments using realistic humanoid avatars (Herrera et al., 2020). However, in artificial agents such as social robots that rely on physical embodiment to interact with the users and the real-world, artificial creative behavior will need to be communicated to users verbally and kinesthetically, namely, using movement as part of their communicative means. To our knowledge, the display of creative behavior *via* physical movement (animacy) in robots is an area requiring further investigation by human–robot interaction researchers.

In human–human interaction (HHI), the study of non-verbal behavior in creative and collaborative tasks has been extensively studied. For instance, Won et al. (2014) suggest that there is a significant correlation between synchronized movements of a pair of humans and creative outcomes. Hence, the display of creative behavior *via* movement is likely to be of high relevance to determine the ultimate value and usefulness of social robots interacting with humans in everyday settings.

The impact of robots' physical presence and their movements has been studied previously across contexts (Vignolo et al., 2017). However, more research is needed to better understand how social robots can effectively use movement in everyday interactions, especially in light of screen-based smart applications and disembodied products that use voice interfaces. What may physical robots offer in terms of functionality and usability that screen and voice agents cannot? What are the design affordances enabled by their physicality and movement possibilities that are not available on screen or voice interaction? The work presented in this article seeks to contribute to the future design of social robots by analyzing how humans perceive the robots' movement when these perform a task that requires creative behavior, such as play.

We propose an adapted scale to measure the perceived creativity in social robots in the context of games and playful activities such as creative storytelling. This scale is inspired in the work of Karwowski et al. (2018) and Karwowski (2014) and is modified here to capture how participants rate the creative skills of robots. We are particularly focused on the domain of creative collaborative interactions that could be eventually implemented using social robots.

The work with social agents as robots is relevant because excessive screen time is associated with ergonomic, visual, and behavioral issues. Furthermore, excessive use of screens for entertainment negatively impacts people at the level of being recently classified as "gaming addiction" (WHO, 2018) and influences people's overall mood, among other effects. At the moment, users intensively use screen-based devices for both work and entertainment services, resulting in extended periods of eye strain and sedentary lifestyle (Aboujaoude and Starcevic, 2015; Alter, 2018; Desmurget, 2020).

This work is an early exploration of HRI that does not rely on a screen to interact with users. We aim to study alternatives to screen and audio interactions using physical robots for interactive creative activities, relying on more natural interactions with artificial agents. We consider that creativity is a central part of HRI due to its importance in the cognitive and social processes involved in playful interactions. Finally, we expect to contribute in the near future in the design of robots encouraging natural, long-term interactions with cognitive and social gains.

2 Background

The design of social robots has shown initial evidence of their potential value for usability and functionality to assist users in everyday life. So far, the main applications of notable market success have been to carpet and floor robot cleaners and toys including robotic pets for therapeutic purposes. For the last 2 decades, researchers and companies have searched for the "killer app" that builds on the affordances of physical robots to transform the lives of users around the world. In that time, screen-based devices and home assistants that use audio interfaces have made substantial gains in market penetration. Currently, there is a need to understand if and how physical robots will be of value for users in their everyday tasks as suggested by the fiction (Miller, 2021). However, to our knowledge, there are no sufficient studies using social robots aiming to study creative interactions between humans and robots, creative performance by the robot, or human's creativity enhancement. In contrast, there are numerous apps, software, and websites for this purpose of screen-based devices.

2.1 Robot's embodiment and movement

Arguably, the most salient affordance that gives social robots an edge over screen and voice assistants is their physical presence. The importance of physicality and movement in communication is evident from "body language" to 4E cognition, that is, the principle that all human cognition is embodied, embedded, enactive, and extended (Lindblom and Alenljung, 2015). In other words, people are not "brains in jars" but rely heavily on their bodies, the physical world around them, other humans, and their contexts to be able to think, communicate, and be fully human.

There have been studies on issues relevant to the design of robots with movement in mind like the study developed by Hoffman and Ju (2014). In the early stages of research in human–robot interaction researchers as Van Breemen (2004) understood that body gestures are a natural channel to communicate robot's agency and social behaviors. Furthermore, the motion of robot agents (mechanic and organic), as one of the main features differentiating robots from AI or computers, has been explored to highlight its relevance from robot esthetic and functional context by Harris and Sharlin (2011). Similarly, Bainbridge et al. (2011) explored how the physical presence of a robot affects human judgments of the robot as a social partner. Looking for the effects of form and motion in robotic agents, Castro-González et al. (2016) studied the attributions of animacy and investigated how the combination of robot's bodily appearance and movement can alter attributions of animacy, likability, trustworthiness, and unpleasantness in the users. Apparently, a Baxter robot executing mechanistic movement was perceived as inanimate. However, the same robot performing naturalistic movements was unpleasant. However, all these previous works have not been placed in the context of creative expressiveness of the robot agents.

Our work aims to explore and understand how animacy plays a role in the perception of social robots in the future design of interactive tasks related to creativity. According to the Oxford English Dictionary, animacy is "...the quality or condition of being alive or animate" (Animacy, 2021) while kinesthetics refers to "...the effort that accompanies a voluntary motion of the body" (Kinaesthesis, 2021). In the context of this research, kinesthetics and animacy refer to the study and perception of body motion, and the kinetic design refers to the use of movement as a design material (Sosa et al., 2015).

2.2 Games as a setup in HRI

We are interested to examine the physicality and moving affordances of social robots (their animacy qualities), and their potential advantages over traditional board games or interactive games such as mobile apps and voice assistants. It may be possible to combine the best of digital and physical affordances to design social robots that can meaningfully augment social games. To this end, studies are needed to assess the impact of the physical presence and movement of robots in these contexts of use.

Social and creative games represent interesting settings to study the interaction with social robots as they create a space for playful semi-structured interactions. In many social games, clear rules exist but significant open-endedness is supported to exercise and enjoy the creativity of oneself and others. Creative games with open rules as Story Cubes have not been used often in HRI. However, games are a popular setup in HRI. For instance, Leite et al. (2009) used chess (to some extent a creative game) as a setup to understand how social presence of robots is perceived. Similarly, interactive storytelling in HRI has been reported as a promising scenario for children's social skills development (Leite et al., 2015, 2017). Storytelling games have the potential to support creative social interactions. We thus select social games, and particularly storytelling games such as the popular "Story Cubes" (Ros and Demiris, 2013; Eladhari et al., 2014; Bae et al., 2016; Gordon and Spierling, 2018) as the site of research for this study.

2.3 Displaying robot creativity

Creativity is an important component of social interaction (Rogers, 1954). Movement has communicative properties that make it an essential part of social interaction between humans (Goldman, 2004), and it is widely regarded as central to embodied experiences including everyday creativity (Svanæs, 2013). Social robots assisting develop creative capacities that have been proposed before demonstrating the importance of physical movement in this type of applications (Hoffman and Ju, 2014). Similarly, storytelling is a creative and social activity that has an entertainment value but is also used to support learning (Sadik, 2008) and health (Plaisant et al., 2000).

Kinesthetic creativity has been studied mostly in artistic performance by humans (Ros and Demiris, 2013; Tan et al., 2018). To our knowledge, this work represents an early approach to the study of kinesthetic creativity in social robots. The research methodology for this study is an experimental design based on previous studies as the ones performed by Salem et al. (2011); Hoffman and Ju (2014); Hoffman et al. (2015); Tung (2016). The effects of robot movement are thus evidenced by their perception of how movement is perceived as a cue, indicating creative agency in social robots.

2.4 Motivation for this exploratory study

In this study, we chose Story Cubes as an experimental setup. This is a game where players create or re-create stories to share with others, thus requiring to some extent creative skills found universally (Brent, 2014). Our study asks whether social robots will have an edge based on their physicality and their basic animacy features to engage in creative activities such as storytelling games. More specifically, it seeks to assess to what extent movement may make a difference in such settings. If the answer to this question is positive, then further work will be needed to identify and evaluate the kinetic principles for the design of playful and creative social robots more specifically. If the answer is negative, then it is more likely that screens and voice interaction devices will be more adequate and arguably even easier to develop, deploy, maintain, and operate in this type of activities rather than robots that use physical bodies to move and reinforce non-verbal interactions. To this end, our study addresses the research gap in how robot movement is perceived when they perform a creative storytelling activity.

In sum, with studies like this, we aim to contribute to the emergent exploration of human–robot creative interaction (HRCI). Thus, here we set to identify the hypothesized connection between the perceived creative agency and the animacy of social robots. Our goal is to evaluate the relevance of robot movements to attribute creativity to social robots. At this stage, we aim to provide a benchmark for future experiments using non-choreographed robot movements. Similarly, we use storytelling games supported by visual cues to study how movement shapes human's perception of creative agency in robots. We need to highlight that most social robots present limitations in terms of dexterity when manipulating objects on a human scale.

3 Methods

3.1 Research goals and questions

This study aims to assess the connection between perceived creative agency and the animacy of social robots. We evaluate the relevance of robot movements to how observers attribute kinesthetic creativity to social robots and the verbal delivery of a story as a creative act. The aim of this experiment is to explore the extent to which robots' movement supports the display of a creative act.

With this exploratory study, we aim to respond the following research questions without proposing any hypotheses due to the complex nature of the interaction and the exploratory nature of the study:

(1) To what extent do humans perceive creative agency in robots when these tell a story as part of a game (display of a creative act)?
(2) To what extent do humans perceive creative agency in robots when these display movements accompanying the delivery of the story?
(3) How are robots perceived as creative agents compared with humans?

3.2 Materials and implementation

3.2.1 The robot

To evaluate the research questions, we programmed a Robotis Mini humanoid robot (See Figure 1A) to play our own version of Story Cubes (storytelling game) in a Wizard of Oz setup. The robot used a female voice in English language (Karen) generated in Mac OS 10.15.7 and played at slow speed (25%) using a bluetooth speaker next to the robot. We chose this robot due its small dimensions for transport and suitability to be used for future experiments using board games. The robot was animated using the Robotis Mini app in iOS 14.4. According to the work of Bernotat et al. (2021) and Kuchenbrandt et al. (2014), the robot design can lead to a male or female perception of the social robot. Hence, we decide to use a female voice to neutralize the possible male perception of the robot considering the very sharp and angular design of the mini humanoid. The robot movements were presented but not choreographed as they were just a stimulus to indicate robot animacy rather than supporting the delivery of the story.

3.2.2 The game

The Story Cubes game is a collaborative board game using six or more dice with adaptable rules and an indeterminate number of players. The goal of the game is to create a story with the contribution of all participants. We designed our Story Cubes for this experiment with a set of six oversized dice that could be visible in the video recordings. After five design iterations using different materials, we used a 30mm, solid white, PDA, 3D-printed dice, which was laser engraved and hand painted. We modified publicly available icons under Creative Commons license for our experiment. Thirty six icons were engraved and six icons were used to reveal "The Ugly Duckling" story for the two different versions used in the experiment.

3.2.3 The story

We use the well-known literary tale titled "The Ugly Duckling" by the Danish author Hans Christian Andersen (Andersen et al., 1995) and a modified version of that story with an alternative ending. Stories are an effective way to connect with other humans and social agents. The creative act of telling a story associated with the movement could possibly lead to significant different perceptions by users of the creative agency of robots. Similarly, a universal story as The Ugly Duckling helps us to reach a significant and diverse pool of participants for this experiment. Finally, the twist plot leading to

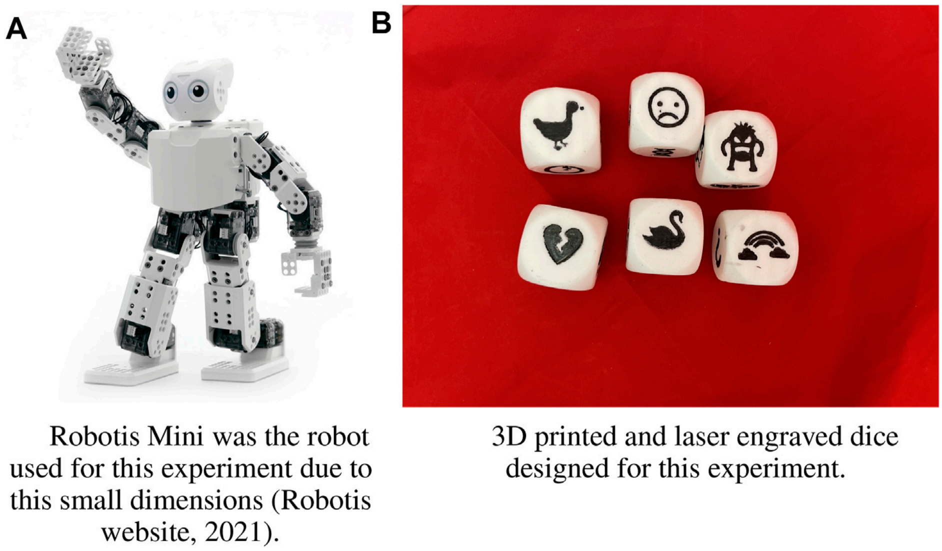


FIGURE 1
Robot and Story Cubes used for this experiment. **(A)** Robot mini was the robot used for this experiment due to this small dimensions (Robotis website, 2021) **(B)** 3D printed and laser engraved dice designed for this experiment.

a different end of the story is a creative act performed by a robot, which could be or not perceived by the user.

3.3 Setup

The standard version of The Ugly Duckling was split in six short sections to match with the six icons shown in Figure 1B and two additional sections for Introduction and Wrap-up. The creative story was deeply discussed by the authors and validated by four experts in creative writing, English literature, and film scripts in its early and latest versions. The latest version uses dinosaur references, which were considered unexpected and perceived as creative without the controversy and the risk of unintentionally offend or hurt feelings of a particular community as the ethics committee suggested (Application HC200985). This last version was validated by one of the specialists in English literature. Both versions, namely, the standard and the creative stories can be read as follows:

Original Story:

-Ok, let us start. Once upon a time [Introduction].
-There was a mama duck who was very surprised with one of her ducklings [1].
-Everyone thought it was very ugly and rejected the poor little duckling [2–3].
-She could not understand why everyone was so cruel only because she was different [3–4].
But a year later, the "ugly duckling" grew into a beautiful swan [5]!
-Then she flew away with a flock of swans and every year returned to say hello to her foster mum [6].
-She, her mum, and siblings celebrate with a happy party around the lake [7].
-The moral of the story is that some people take longer to develop and find their true beauty [Wrap-up and Robot Reflection].

Creative Story (The bold font indicates the creative twist vs. the original story):

-Ok, let us start. Once upon a time [Introduction].
-There was a mama duck who was very surprised with one of her ducklings [1].
-Everyone thought it was very ugly and rejected the poor little duckling [2–3].
-She could not understand why everyone was so cruel only because she was different [3–4].
-But a year later, the 'ugly duckling' grew into a beautiful flying dinosaur [5.1]!
-A pterosaurs called Quetzalcoatlus [5.2].
-She discovered that she was different, not ugly [5.3].
-Then she flew away with a flock of happy pterosaurs and every year returned to say hello to her foster mum [5.4].
-Then she flew away with a flock of swans and every year returned to say hello to her foster mum [6].
-She, her mum, and siblings celebrate with a happy party around the lake [7].
-The moral of the story is that some people take longer to develop and find their true beauty [Wrap-up and Robot Reflection].

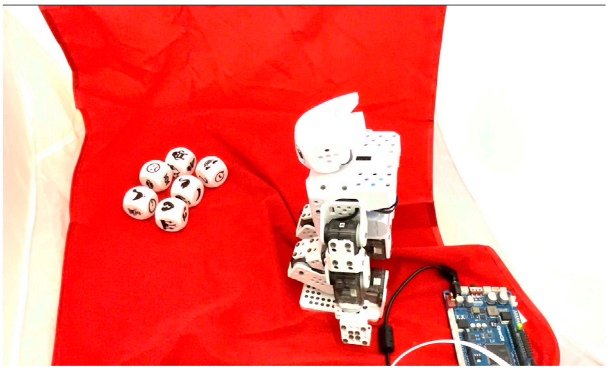


FIGURE 2
Snapshot of the video showed to the participants. The setup used under the four different conditions of the experiment is the same (cubes, robot, and microcontroller in the shot).

The robot, the dice, and the speaker (not visible) were allocated in a photo tent in order to isolate them from the external stimuli, avoid distractions for the users and make possible the replication of the experiment. A high contrast, red carpet was used to highlight the dice and the robot. After several tests, we decided to use the video, which was recorded in from the top-left corner of the photo tent using an iPhone 12 mini with 0.8x digital zoom. This setup allows to the viewers to have a full view of the robot and dice. We consider that using this point of view allows the user to understand the creation of the story by the robot. See Figure 2. The length of the videos was shorter when the standard story was displayed due to the three extra sentences used in the creative story. These three extra sentences in the creative story aim to reinforce the novel character of the story and assure that the participant noticed the plot twist. This setup was implemented by the suggestions of the experts in writing and literature.

Four videos were displayed to the participants showing two different robots performing a creative task. The creative task consists of storytelling performed by the robot. The method to tell the story is supported by visual cues in the form of icons on dice and the robot manipulating them. Once the bowl covering the dice was removed, the robot started one of the proposed stories with the movements depending on the conditions.

In sum, we programmed a humanoid robot to tell the original version of the "The Ugly Duckling" and a modified version of the same story (creative story) under the experimental condition using cubes with visual imagery such as those used in the "Story Cubes" creative game. Under the control condition, the robot remains static while telling the story, while under the experimental condition the robot performs movements to accompany telling the story kinesthetically. These interactions were video recorded, and participants were requested to fill a survey evaluating the creative agency of the robot.

4 Experimental design

We designed a 2 × 2 between-subject online study. The factors are the story and robot movements. The robot can tell the standard story or the creative story (dinosaur plot twist), and the robot can display movements or not (still) during the storytelling. Hence, we test four conditions displayed by the robot: still robot and standard story (SS), still robot and creative story (SC), moving robot and standard story (MS), and moving robot and creative story (MC). As a between-subject study, the participants of this online study were exposed to one of the conditions mentioned earlier, that is, a participant under the condition MC would see a robot gesticulating and telling The Ugly Duckling story with the plot twist of the dinosaur. The dice and videos of this experiment can be requested contacting the corresponding author. The videos are unlisted in YouTube but can be reviewed using the next links: Video MS condition, Video MC condition, Video SC condition, and Video SS condition. See Table 1.

4.1 Online survey and measurements

This study was approved by the human ethics committee of the University of New South Wales, application HC200985 reviewed by the HREAP Executive. Similarly, this study was funded using the Scientia Fellowship (PS-46183) development package provided by the UNSW. We implemented the survey on Qualtrics licensed for the UNSW. A survey flow was created to assign the participants under the four conditions randomly, see Figure 3. The survey was distributed using prolific.co. The survey was designed as follows: First, the participant information and then the consent form. Once the participant agreed to participate in the studio, he/she was directed to the next page, and demographic information and confirmation of the Prolific ID was collected. The identity of the participants is anonymous, and demographic information was collected such as age, gender, occupation, location, and level of education.

Once the demographic information was collected, the Short Scale of Creative Self (Karwowski et al., 2018) questionnaire was applied. The questionnaire is confirmed by eleven questions measuring creative self-efficacy (CSE) and creative personal identity (CPI). The questions are as follow:

(1) I think I am a creative person (CPI).
(2) My creativity is important to who I am (CPI).
(3) I know I can efficiently solve even complicated problems (CSE).
(4) I trust my creative abilities (CSE).
(5) Compared to my friends, I am distinguished by my imagination and ingenuity (CSE).

TABLE 1 Four experimental conditions. Each condition shows the factorial combination displaying a particular performance of the robot.

Experimental design	Standard story (control condition)	Creative story (experimental condition)
Still robot	Still standard condition (SS)	Still creative condition (SC)
Moving robot	Moving standard control (MS)	Moving creative condition (MC)

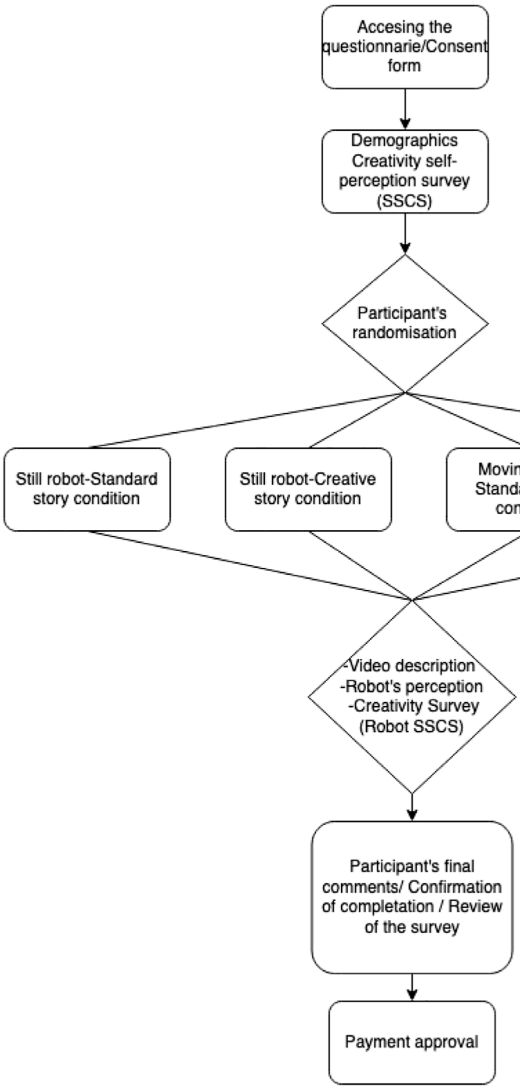


FIGURE 3
Experimental procedure. We applied the questionnaire using Qualtrics. We distributed it using Prolific.co.

(6) Many times I have proven that I can cope with difficult situations (CSE).
(7) Being a creative person is important to me (CPI).
(8) I am sure I can deal with problems requiring creative thinking (CSE).
(9) I am good at proposing original solutions to problems (CSE).
(10) Creativity is an important part of me (CPI).
(11) Ingenuity is a characteristic which is important to me (CPI).

Questions one, two, seven, ten, and eleven gauge CPI and questions three, four, five, six, eight, and nine are assigned to the CSE. Following the questions, the participants watched one of the four videos. We confirmed that the video was watched *via* a check button and requesting a brief description of the video by the participant. Next, the two first authors proposed a modified version of the SSCS to be applied to robots playing creative task. In this case, the task is framed in the playing of our version of Story Cubes. As the original SSCS, our Likert scale goes from one to five. The anchors are definitely not 1) and definitely yes (5). We will validate this adapted scale in a following article. The questions proposed are listed as follow.

(1) I think that the way the robot played Story Cubes shows it is a creative robot.
(2) The creativity of the robot playing Story Cubes is important to how it behaves.
(3) The robot efficiently plays Story Cubes.
(4) I trust the robot's creative abilities to play Story Cubes.
(5) Compared to other players of Story Cubes, the robot is distinguished by its imagination and ingenuity.
(6) I think that the robot can consistently create good stories when playing Story Cubes.
(7) Being creative is important for a player of Story Cubes.
(8) The robot can deal with problems requiring creative thinking.
(9) The robot is good at proposing original stories in Story Cubes.
(10) Creativity is an important part of how the robot plays Story Cubes.
(11) Ingenuity is a characteristic which is important to how the robot plays Story Cubes.
(12) I think that the robot can consistently create good stories when playing Story Cubes. Additional question:

In addition, we added a 12th question. The aim of this question is to summarize the general impression of the participant in just one score. The Godspeed questionnaire was applied after the SSCS and general impressions over the study were requested to the participant. The survey concludes confirming the end and providing a code that will allow the compensation for the participant by Prolific. The content of the survey is available by request to the corresponding author.

4.2 Participants

A total of 297 participants were recruited using the platform Prolific. All participants were over 18 years, and there were no restrictions on gender, formal education, income, or other demographics. Due to technical issues, just 242 participants were exposed to one of the four conditions and compensated by completing the questionnaire. We used the data of 239 participants as the SSCS score as three of their data were not recorded. Participants were paid the equivalent of 1.27 British pounds (or 2.2 Australian dollars) for their participation. The first half of the study was run in February 2021, and the second part was 1 week later. One fourty five participants were male, 89 female, three non-binary, and two did not specify. The average age was 26.9 years (SD = 8.58). Participants came from a range of locations; 38.5% from North America (Canada, United States, and Mexico), 34.7% from Europe, 13% from the United Kingdom, 7.5% from South America, 4.2% from Oceania, and 2.1% from Africa. The education levels are distributed as follows: 42.7% university degree, 38.5% high school, 9.6% masters, 1.7% PhD, 5.4% vocational education, 1.7% primary education, and 0.4% other. In terms of occupation, 42.3% were students, 38.9% employed (6.3% IT and software, 3.3% artist, 2.5% freelance, 1.3% researcher, and 25.5% other), 12.1% were unemployed, 2.5% homemakers, 0.4% retired, and 1.3% did not specify.

The average time to fill the survey was 9 min 20 s. We suggested to the participants to use a device with a large screen to have a better user experience and show the video and survey consistently; 72.8% used Windows 10, 8.4% used Macintosh, and 18.8% used other platform (Android 10 5.4%, iPhone 4.2%, Windows 6.1 3.8%, Windows 6.3 1.7%, Android 11 1.3%, Android 9 1.3%, Linux x86 0.8%, and Ubuntu 0.4%).

We aimed to allocate at least 60 participants per condition. We did not fully record data as three subjects lost in the SSCS score. Participants were randomly allocated as follows: (SS) = 62 participants (minus two missed SSCS scores), SC = 61 participants, MS = 61 participants (minus one missed SSCS score), and MC = 58 participants. The average human SSCS score was 3.89 (SD = 0.66), no significant differences were found among the different experimental conditions (ANOVA). Participant's results are stored in a standard online spreadsheet, and the statistical analysis was made using IBM SPSS.

5 Results

In order to address the exploratory questions of this study, we performed multiple 2 × 2 factorial analyses of variance (ANOVA); the factors were the story (standard vs. creative) and movements (still vs. moving). We defined seven dependent variables: the SSCS, the question twelve, and the five Godspeed items (anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety). In addition, we performed a Pearson's correlation present among the humans and robots CPI, CSE, and SSCS scores to check the internal consistency and possible human–robot significant correlations among the scores.

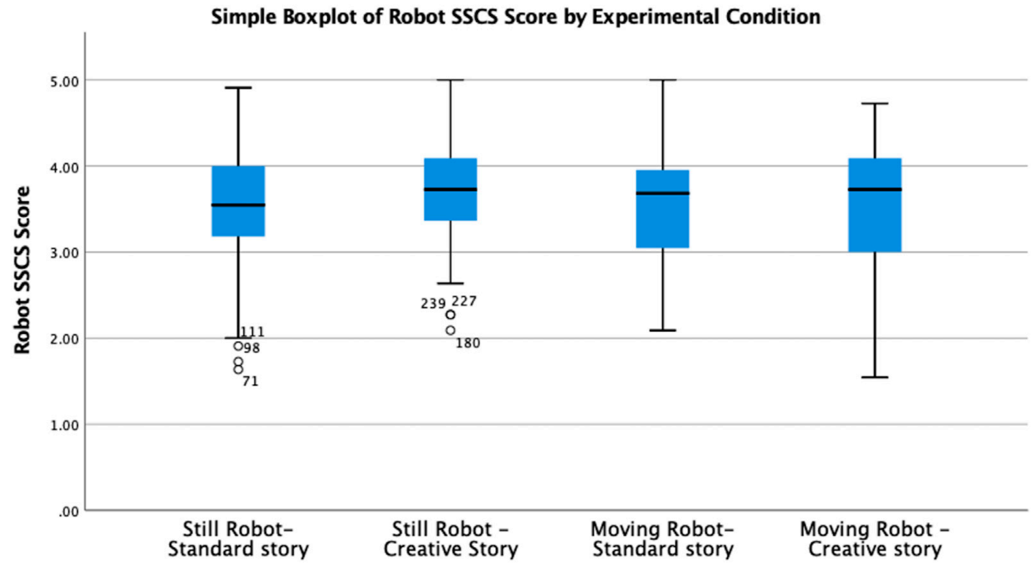


FIGURE 4
Boxplot per group of the SSCS applied to the robot with outliers. No significant differences were found among the means of the groups.

5.1 Short Scale of Creative Self (SSCS) score applied to the robots

This variable is our main indicator to assess how participants perceive robots as creative agents. The average SSCS score of the robots under the four different conditions are as follows: SS = 3.48 (SD = 0.74), SC = 3.66 (SD = 0.61), MS = 3.58 (SD = 0.68), and MC = 3.53 (SD = 0.80). As the original SSCS, our Likert scale goes from one to five. The anchors are definitely not 1) to definitely yes (5).

We ran a two-way ANOVA with the SSCS score as a dependent variable and the story and movements as factors. Residual analysis was performed to test for the assumptions of the two-way ANOVA. The assumption of homogeneity of variances was not violated, as assessed by Levene's test for equality of variances, $p = 0.122$. Data were normally distributed as assessed by the Kolmogorov–Smirnov test ($p = 0.200$). There were six outliers as assessed as being located less than 3 box-lengths from the edge of the box in a boxplot. They were not removed from the following ANOVA and showed that they did not affect the results even without them. Neither significant main effects nor interaction effects were found. See Figure 4. Similarly, pairwise comparisons were run aiming to find simple main effects. However, non-significant effects were found again.

Hence, as suggested in Laerd (2021), a further robust ANOVA was run. We used SigmaPlot to run a Kruskal–Wallis non-parametric ANOVA. Once again, neither main effect nor interaction effect was found. KW = 2.11, df = 3, and $p = 0.550$.

5.2 Question twelve and Godspeed items

A similar procedure was followed for the analysis of the rest of the dependent variables. The effects were found only for the items of animacy, likeability, and perceived safety. Data were normally distributed, as assessed by the Kolmogorov–Smirnov test for animacy ($p = 0.200$) and likeability ($p = 0.200$) but not for perceived safety ($p < 0.001$) and Q12 ($p < 0.001$). The assumption of homogeneity of variances was not violated for animacy (0.989), perceived safety ($p = 0.173$), and Q12 ($p = 0.909$) as assessed by Levene's test for equality of variances. However, it failed for likeability ($p = 0.006$).

We decided to maintain the outliers assessed as those being located less than 3 box-lengths from the edge of the box in a boxplot. The results indicate that movement has a main effect on how participants perceive the robots under the different conditions. Animacy $F(3,235) = 39.777$, $p < 0.001$, likeability $F(3,235) = 12.824$, $p < 0.001$, and perceived safety of the robot $F(3,235) = 19.127$, $p < 0.001$. Due to the violation of the assumption of homogeneity for likeability, a robust non-parametric Kruskal–Wallis ANOVA was run for this variable. Movement has a significant main effect. KW = 10.205, df = 3, and $p = 0.017$. See Table 2.

5.3 Person's correlations among CPI, CSE, and SSCS in humans and robots

A Pearson's 1-tailed correlation was carried out to assess the relationship between the human CPI, CSE, and SSCS scores and

TABLE 2 Bold font means this item had a significant main effect on the movement condition.

Means and SD for Question 12 and Godspeed items under the movement main effects condition	Still robot standard story	Still robot creative story	Moving robot standard story	Moving robot creative story
Q12	3.27 (1.06)	3.25 (1.06)	3.57 (0.98)	3.45 (1.08)
Anthropomorphism	2.16 (0.73)	1.99 (0.77)	2.27 (0.76)	2.15 (0.76)
Animacy	2.36 (0.76)	2.33(0.74)	3(0.68)	2.84(0.66)
Likeability	3.49(0.96)	3.61(0.96)	3.96(0.62)	3.89(0.73)
Perceived intelligence	3.48 (0.87)	3.47 (0.75)	3.67 (0.70)	3.45 (0.72)
Perceived safety	2.88(0.61)	2.74(0.37)	3.12(0.54)	3.10(0.56)

TABLE 3 Pearson's correlations among the CPI, CSE, and SSCS scores. The scores are internally consistent.

Pearson's Correlations among the CPI, CSE, and SSCS scores for humans and robots (p <0 .001)	Human CSE	Human CPI	Human SSCS	Robot CSE	Robot CPI	Robot SSCS
Human CSE	1					
Human CPI	0.573	1				
Human SSCS	0.883	0.891	1			
Robot CSE				1		
Robot CPI				0.673	1	
Robot SSCS				0.931	0.896	1

similar scores granted to the robot. We aimed to find if correlations between people perceiving themselves as highly creative project this on the robot's creative act. The scores of the 239 participants were analyzed. There were statistically significant, moderate, and strong positive correlations between CPI–CSE, SSCS–CSE, and SSCS–CPI in humans, and similar pattern of correlations among the robot scores suggesting consistency in our proposed scale and its internal scores. However, no significant correlation appeared between human scores and robot scores. We should highlight that overall the means of the human scores were higher than that of the robot scores under all the experimental conditions and all the scores. See Table 3.

6 Discussion

This study aims to explore the design of future social robots using a quantitative approach and qualitative insights from the users. With this study, we want to encourage a discussion in the domain of robot's perceived creativity and explore robot movement supporting the delivery of a creative act. However, studying creativity using quantitative approaches presents a number of challenges. We frame our findings in the context of games as a mean to sustain long-term, meaningful human–robot creative interactions and several considerations should be taken.

To answer the first research question: to what extent do humans perceive creative agency in robots when they tell a story as part of a game (display of a creative act)? We assured that participants watched the video requesting a checkbox validation and a brief description of what they saw. Participants frequently used the word "story" or even specifically "ugly duckling story" in these descriptions; 92.5% of the participants mentioned the word "story" or similar (tale, history, story line, and fable) when describing the video. In few cases typos were present and derivations such as history, story, or study were used but the intention was taken as the same. Furthermore, 100% of the participants referred directly or indirectly to the act of telling a story even when they did not use the word in their descriptions. For instance, "it is about a robot that tells ugly duck that turns into swan," "the process to find the real beauty," or "It is about a robot that tells ugly duck that turns into swan."

We highlight the fact that the SSCS scores of the robots, even when they are not significantly different, were above the mean (2.5) of the 1–5 scale of the SSCS score as indicated in section 5.1. Even though the robot scores were lower that the human scores, the minimal score was above 3.4 for the still robot standard story condition. The scores per condition were: SS = 3.48 (SD = 0.73), SC = 3.66 (SD = 0.61), MS = 3.58 (SD = 0.68), and MC = 3.53 (SD = 0.80). Hence, we can claim that participants were aware that the robot was performing a creative act delivering a story. Future work can involve further statistical analysis comparing with a specific benchmark, indicating a minimal score that a social agent considered for a creative agent. Similarly, a face-to-face setup could be more appropriate to perform an experiment of this nature since it is likely that the robot embodiment has a significant impact in the participant's perceptions compared to virtual agents.

For the second question: to what extent do humans perceive creative agency in robots when these display movements accompanying the delivery of the story? We considered that robot movement would be a variable moderating the effect of the story on how people perceive robots. The marginal means graph can wrongly lead to conclude that movements moderate the SSCS score. However, when inspecting the boxplot, it is clear that means among all the experimental conditions are not significant. See 4. Although we did not notice main or interaction effects in the SSCS score, we did notice significant effects in three of the items of the Godspeed scale. These items are animacy, likeability, and perceived safety.

In the case of animacy, we observed that participants in this study could notice the movements of the robot as they score significantly highly in animacy to the moving robots independently of which story the robot is telling. This shows that participants are aware of the movement and how the movement impacts participant's perceptions in terms of likeability as they rank moving robots MS = 3 (0.68) and MC = 2.84 (0.66) significantly higher than still robots SS = 2.36 (0.76) and SC = 2.33 (0.74). See Table 2.

The robot movements were not mentioned frequently in the description of the video as the story. However, some participants used anthropomorphic terms, that is, "a creepy robot gives its version of The Ugly Duckling by Hans Christian Andersen. It also does a decent MC hammer impression."...The movements of the robot were however a bit erratic and did not match that well with the story it was telling. "In a way I feel like I was arranged to tell that story but I liked the movements and the appearance of the robot."

In terms of likeability, as shown in Table 2, participants scored significantly higher to the moving robots in terms of likeability. MS = 3.96 (0.62) and MC = 3.096 (0.62) significantly higher than still robots SS = 3.49 (0.96) and SC = 3.61 (0.96). The significant main effect of the movements in likeability is aligned with previous studies using games (Sandoval et al., 2016b; a, 2020). Apparently, humanoid robots tend to be likeable when they perform unexpected tasks that can be interpreted as social or creative. Future studies could test other robot embodiment perceived as less humanoid. An illustrative comment in how some participants perceive the robot was: "how this robot looks like and how it feels like compared to a human being. By his movements, he looked really happy, but by talking only sometimes we cannot understand how smart a robot can be. He was very smart, and we can definitely see that his voice was not recorded at some point."

At the beginning of this experiment we considered that the factor of movement would be a moderating variable, supporting the delivery of the story by the robot. In other words, robot movement would lead to higher scores for the robots in both kinds of stories or at least in one of the stories. However, evidence for this were not registered, a possible reason for that is the type of the robot's movement. For this experiment, we intentionally designed a set of robot movements that are part of the standard programming of the robot but are not synchronized with the delivery of the story. The reason for this is that in future applications using robots for social games, it is unlikely that robot movements will always be customized to their dialogs. Better synchronized and choreographed movements would be an obvious stimulus that could cause significant main effects in the SSCS score and how people perceive robots as creative agents. This type of movements can be tested in future studies. In terms of perceived safety, moving robots were perceived as safer than still robots. This is a result worth considering further, especially taking into account that this study is an online experiment and not a face-to-face setup.

Question twelve in the survey: "I think that the robot can consistently create good stories when playing Story Cubes" was added to allow the participants to summarize their impressions from the previous SSCS questions. The movement main effect was not significant ($p = 0.64$) when the participants answered this question. However, this provides an intriguing result to be explored in future studies as the comments of the participants suggest. Certainly, the participants perceived the movement of the robot but not necessarily the novelty of the story. Independently of the story, participants rated slightly higher for the moving robots than still robots as agents, which can create good stories when playing Story Cubes. Even as a marginal result, this finding points to the importance to study a range of robot movement approaches in future work. Further qualitative analysis is required to explore.

For question three in the survey: how are robots perceived as creative agents compared with humans? In all the experimental conditions robots scored lower than humans for the SSCS scores. Robot scores can be seen in Section 5.1 and human SSCS scores are as follow SS = 3.88(SD = 0.68), SC = 3.85 (SD = 0.83), MS = 3.88 (SD = 0.70), and MC = 3.95 (SD = 0.66). Pearson correlation does not suggest any significant correlation among the human and robot scores. However, there are significant correlations among the sub-scores CPI and SCE in both human and robot scores, which could suggest that the internal consistency of the original scale and our adaptation for assessing robot creativity have potential as measurement tools for future studies in human–robot creative interactions. See Table 3.

Some of the participants' comments suggest that a further exploration of this question could be relevant as they were enthusiastic about the capabilities of the robots comparing with previous human creative experiences. To illustrate, first of all, the participants' adored the robot. they thought it was cute plus they were impressed with its abilities, too! they meant they can probably kind of guess how it works, but still exclaimed that it was just mind-blowing! The participants' loved that the message of the story told by the robot was so wholesome! As a musician and somewhat a songwriter, they find it astonishing that how it can come up with a good story in such a quick amount of time, and they wished keep it up. As for the study itself, they liked the way the text lighted up when the mouse hovered over it, they have not seen it a lot, if anytime. In addition, they had to check up on two of the English words used in the study, which they highly value as an educational feature. They thoroughly enjoyed the experience. "This was an interesting concept to consider, and the participants' would honestly like to see more content involving AI and Story Cubes." They are impressed how robot can tell people a story based on random images.

7 Conclusion

Creativity can be considered an aspect of autonomy and agency in social agents that is different from intelligence, logic, and strategy. The current understanding on how creativity is displayed by robots is still limited. This study aimed to inform the design possibilities of human–robot creative interaction (HRCI) and provides a reference for future studies exploring the main factors involved in the creative interaction between humans and robots. Our findings show that the setup used in this study does not trigger higher scores in the SSCS, differentiating the robots as creative agents. However, movement does show main effects in the scores of animacy, likeability, and perceived safety of the Godspeed scale. Furthermore, the scores of the moving robots were above the media in all the cases (although they were lower than the SSCS scores in humans). In terms of how robots are perceived as good storytellers (question 12 in the robot SSCS), even when the scores are not significantly different, the results provide an important insight in how to continue the development of future experimental studies, for instance, the need to perform similar study in a face-to-face setup and using other robot embodiments beyond humanoid robots.

The study of creativity in robots shows a research gap when addressed in playful, creative, and collaborative activities such as board games. We chose a playful task (a storytelling game) to empirically evaluate the extent to which a robot's physical embodiment may cause humans to attribute creative agency to a robot. The Story Cubes game offers a means to further assess the display of creativity in robots considering the applicability of this setup as entertainment and for the development of cognitive skills, spacial memory, decision making, and collaborative skills (Wu et al., 2012; Unbehaun et al., 2019).

Furthermore, we consider that our approach is useful in aiming strategies for long-term interaction and as an alternative to avoid screen addiction and contribute to a better mental health in the digital age (Aboujaoude and Starcevic, 2015; Sandoval, 2019). Even in the Story Cubes mobile app, the user experience is visibly compromised when compared to the physical cubes. Considering this, we highlight the importance of perceived creativity in social robots to further develop the early work in artificial creativity. It looks like it is critical to explore advantages and disadvantages among robotic interfaces that display and support creative interactions. To this end, when users are exposed to stimuli related to creative robots it seems critical to set their expectations in this type of studies. One participant said, for instance: "the study itself was fine. The premise, however, is something that has not been fully explained." For example, is this a prototype of a children's toy? Is it a learning device? Is it a diagnostic tool?" Finally, one of the comments of the participants is encouraging to continue the development of studies in creative robots playing storytelling games. Assuming the robot really came up with the story using its own creativity and it was not programmed into it, the participant was very impressed with the level of depth which the story had. In that sense, the robot could even be more creative than a lot of humans. The participant also thinks that its use of words and its manner of speech is properly human-like, that is not to say, we could not feel the robotic nature behind it at all. In my opinion, some work should be performed on the robot's movement, and how it connects to whatever it is saying in a way that makes more sense. The participant wished us good luck with the study and with further development of creative robots."

7.1 Limitations and future work

The main technical limitations to implement a robot board game have been discussed before by Sandoval et al. (2021). The game Story Cubes in particular has a significant element of improvisation and randomness, and the translation of the cubes to create a consistent story is a technical challenge for robots that genuinely synthesize stories from these stimuli. Similarly, the vision system required to read the cubes accurately under a range of lighting conditions and angles would require significant technical work. We are currently working on an implementation where participants and robots play Story Cubes in a shared physical setup for a future study. Choreographic movements of the robots will be programmed and displayed on a bottom-up strategy that starts by incorporating movements in an increasing level of sophistication and detail. Then, we will compare with the random movements of this experiment. We consider that this would inform robot designers to incorporate movement that achieves a balance between creating a meaningful human–robot creative interaction (HRCI) while drawing from a library of gestures, postures, and body movements suitable for fluid communication. In terms of data collection and analysis, we plan to conduct a thematic analysis of participants' comments. Furthermore, future experimental designs could include different robot embodiments (humanoid vs. non-humanoid), variants of Story Cube games, and different stories (original stories vs. well-known stories). Finally, a more exhaustive validation of our version of the SSCS (performing factorial and reliability analysis) may be required to assess future interaction face-to-face setup in a more robust manner.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by UNSW ethics committee, application HC200985 reviewed and approved by the HREAP Executive. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author Contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This project was funded by the UNSW Scientia Fellowship PS-46183-A. This fund pays for the open access publication fees, equipment, participant's remuneration, and the rest of the research expenses associated with this study.

Acknowledgments

Sandoval wants to thank to Prof Sosa for all the intense and highly collaborative discussions that lead to this article and the rest of the authors for their valuable contributions. Also thanks to Prof Alfredo Jimenez (IPN-CEPROBI) for his advice on selecting and running additional statistical analysis. Thanks to Karam Hussain at the UNSW Makerspace Paddington campus for his support in the development of the Story Cubes used in the experiments. Also thanks to the story reviewers Carolina Posadas, Jennifer Breakelaar, and Miranda Verswijvelen. Sandoval wants to thank Sebastian for inspiring him to finish the first draft of this manuscript one day before his birth.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2022.695162/full#supplementary-material

References

Aboujaoude, E., and Starcevic, V. (Editors) (2015). *Mental health in the digital age: Grave dangers*. illustrated edition edn (Great Promise: Oxford University Press).

Alter, A. (2018). *Irresistible: The rise of addictive technology and the business of keeping us hooked*. reprint edition edn. USA: Penguin Books.

Alves-Oliveira, P., Arriaga, P., Cronin, M. A., and Paiva, A. (2020). "Creativity encounters between children and robots," in Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction (New York, NY, United States: Association for Computing Machinery), 379–388.

Alves-Oliveira, P., Arriaga, P., Paiva, A., and Hoffman, G. (2017). "YOLO, a robot for creativity: A Co-design study with children," in Idc '17: Proceedings of the 2017 conference on interaction design and children (New York, NY, USA: Association for Computing Machinery), 423–429. doi:10.1145/3078072.3084304

Andersen, H. C., Letts, K., and Ello, C. (1995). *The ugly duckling*. London, England: I. E. Clark Publications.

Animacy (2021). *Animacy, n.* Oxford University Press.

Bae, B.-C., Seo, G., and Cheong, Y.-G. (2016). "Towards procedural game story creation via designing story cubes," in International conference on interactive digital storytelling (Springer), 399–402.

Bainbridge, W. A., Hart, J. W., Kim, E. S., and Scassellati, B. (2011). The benefits of interactions with physically present robots over video-displayed agents. *Int. J. Soc. Robot.* 3, 41–52. doi:10.1007/s12369-010-0082-7

Bernotat, J., Eyssel, F., and Sachse, J. (2021). The (Fe)male robot: How robot body shape impacts first impressions and trust towards robots. *Int. J. Soc. Robot.* 13, 477–489. doi:10.1007/s12369-019-00562-7

Brent (2014). *Creative story telling with story cubes*. London, UK: EdTech.tv.

Castro-González, l., Admoni, H., and Scassellati, B. (2016). Effects of form and motion on judgments of social robots animacy, likability, trustworthiness and unpleasantness. *Int. J. Human-Computer Stud.* 90, 27–38. doi:10.1016/j.ijhcs.2016.02.004

Desmurget, M. (2020). *La fábrica de cretinos digitales: Los peligros de las pantallas para nuestros hijos*. Madrid: Ediciones Península.

Digital Humans, Uneeq Ltd (2021). *Digital humans: Conversational AI solutions beyond just chatbots*. Auckland, New Zealand: Ponsonby.

Eladhari, M. P., Lopes, P. L., and Yannakakis, G. N. (2014). "Interweaving story coherence and player creativity through story-making games," in International conference on interactive digital storytelling (Springer), 73–80.

Gizzi, E., Castro, M. G., and Sinapov, J. (2019). "Creative problem solving by robots using action primitive discovery," in 2019 joint IEEE 9th international conference on development and learning and epigenetic robotics (ICDL-EpiRob) (IEEE), 228–233.

Goldman, E. (2004). *As others see us: Body movement and the art of successful communication*. New York: Psychology Press. Google-Books-ID: mGlKk5sQOVgC.

Gordon, A. S., and Spierling, U. (2018). "Playing story creation games with logical abduction," in International conference on interactive digital storytelling (Springer), 478–482.

Harris, J., and Sharlin, E. (2011). "Exploring the affect of abstract motion in social human-robot interaction," in *2011 ro-man* (Atlanta, GA, USA: IEEE), 441–448. doi:10.1109/ROMAN.2011.6005254

Herrera, F., Oh, S. Y., and Bailenson, J. N. (2020). Effect of behavioral realism on social interactions inside collaborative virtual environments. *Presence. (Camb).* 27, 163–182. doi:10.1162/pres_a_00324

Hoffman, G., Forlizzi, J., Ayal, S., Steinfeld, A., Antanitis, J., Hochman, G., et al. (2015). "Robot presence and human honesty: Experimental evidence," in 2015 10th ACM/IEEE international conference on human-robot interaction (HRI) (Portland, OR, USA: IEEE), 181–188.

Hoffman, G., and Ju, W. (2014). Designing robots with movement in mind. *J. Hum. Robot. Interact.* 3, 89. doi:10.5898/JHRI.3.1.Hoffman

Karwowski, M. (2014). Creative mindsets: Measurement, correlates, consequences. *Psychol. Aesthet. Creativity, Arts* 8, 62–70. doi:10.1037/a0034898

Karwowski, M., Lebuda, I., and Wiśniewska, E. (2018). Measuring creative self-efficacy and creative personal identity. *Int. J. Creativity Problem Solving* 28, 45–57. Place: Republic of Korea Publisher: Korean Assn for Thinking Development.

Kinaesthesis (2021). *Kinaesthesis, n.* Oxford University Press.

Kuchenbrandt, D., Häring, M., Eichberg, J., Eyssel, F., and André, E. (2014). Keep an eye on the task! How gender typicality of tasks influence human–robot interactions. *Int. J. Soc. Robot.* 6, 417–427. doi:10.1007/s12369-014-0244-0

Laerd, S. (2021). *Two-way ANOVA in SPSS statistics*. London, UK: Laerd Statistics Premium.

Leite, I., Martinho, C., Pereira, A., and Paiva, A. (2009). "As Time goes by: Long-term evaluation of social presence in robotic companions," in RO-MAN 2009 - the 18th IEEE international symposium on robot and human interactive communication (Toyama, Japan: IEEE), 669–674. doi:10.1109/ROMAN.2009.5326256

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., et al. (2015). "Emotional storytelling in the classroom: Individual versus group interaction between children and robots," in Hri '15: Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction (New York, NY, USA: Association for Computing Machinery), 75–82. doi:10.1145/2696454.2696481

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., et al. (2017). Narratives with robots: The impact of interaction context and individual differences on story recall and emotional understanding. *Front. Robot. AI* 4. doi:10.3389/frobt.2017.00029

Lindblom, J., and Alenljung, B. (2015). "Socially embodied human-robot interaction: Addressing human emotions with theories of embodied cognition," in *Handbook of research on synthesizing human emotion in intelligent systems and robotics (IGI global)* (Pennsylvania, USA: IGI Global), 169–190.

Miller, M. N. (2021). *Love, Death and Robots season 2, episode 1 recap*. Scotts Valley, California: Automated Customer Service.

Myoo, S. (2019). Creative robots. *Studia de Arte Educ.* 14, 30–39. doi:10.24917/20813325.14.3

OpenAI (2021). *OpenAI*. San Francisco: OpenAI.

Pham, T. X. N., Hayashi, K., Becker-Asano, C., Lacher, S., and Mizuuchi, I. (2017). "Evaluating the usability and users' acceptance of a kitchen assistant robot in household environment," in 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN) (Lisbon, Portugal: IEEE), 987–992. doi:10.1109/ROMAN.2017.8172423

Plaisant, C., Druin, A., Lathan, C., Dakhane, K., Edwards, K., Vice, J. M., et al. (2000). "A storytelling robot for pediatric rehabilitation," in Assets '00: Proceedings of the fourth international ACM conference on Assistive technologies (New York, NY, USA: Association for Computing Machinery), 50–55. doi:10.1145/354324.354338

Rogers, C. R. (1954). Toward A theory of creativity. *ETC A Rev. General Semant.* 11, 249–260.

Ros, R., and Demiris, Y. (2013). "Creative dance: An approach for social interaction between robots and children," in *International workshop on human behavior understanding* (Springer), 40–51.

Sadik, A. (2008). Digital storytelling: A meaningful technology-integrated approach for engaged student learning. *Educ. Technol. Res. Dev.* 56, 487–506. doi:10.1007/s11423-008-9091-8

Salem, M., Rohlfing, K., Kopp, S., and Joublin, F. (2011). "A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction," in *2011 ro-man* (Atlanta, GA, USA: IEEE), 247–252. doi:10.1109/ROMAN.2011.6005285

Sandoval, E. B. (2019). "Addiction to social robots: A research proposal," in 2019 14th ACM/IEEE international conference on human-robot interaction (HRI) (Daegu, Korea (South): IEEE), 526–527. doi:10.1109/HRI.2019.8673143

Sandoval, E. B., Brandstatter, J., Yalcin, U., and Bartneck, C. (2020). Robot likeability and reciprocity in human robot interaction: Using ultimatum game to determinate reciprocal likeable robot strategies. *Int. J. Soc. Robot.* 13, 851–862. doi:10.1007/s12369-020-00658-5

Sandoval, E. B., Brandstetter, J., and Bartneck, C. (2016a). "Can a robot bribe a human? The measurement of the negative side of reciprocity in human robot interaction," in 2016 11th ACM/IEEE international conference on human-robot interaction (HRI) (Christchurch, New Zealand: IEEE).

Sandoval, E. B., Brandstetter, J., Obaid, M., and Bartneck, C. (2016b). Reciprocity in human-robot interaction: A quantitative approach through the prisoner's dilemma and the ultimatum game. *Int. J. Soc. Robot.* 8, 303–317. doi:10.1007/s12369-015-0323-x

Sandoval, E. B., Shi, J., Cruz-Sandoval, D., Li, B., Cappuccio, M., and Rosenbaum, S. (2021). "A prototype of a robot memory game: Exploring the technical limitations of human-robot interaction in a playful context," in Hri '21: Companion of the 2021 ACM/IEEE international conference on human-robot interaction (New York, NY, USA: Association for Computing Machinery), 195–199. doi:10.1145/3434074.3447158

Saunders, R., Chee, E., and Gemeinboeck, P. (2013). "Evaluating human-robot interaction with embodied creative systems," in Proceedings of the fourth international conference on computational creativity (Sydney, Australia: IEEE), 205–209.

Sosa, R., Lee, J. B., Albarran, D., and Otto, K. (2015). "From concept to specification maintaining early design intent," in *ICoRD'15 – research into design across boundaries volume 2*. Editor A. Chakrabarti (New Delhi, India: Springer, Smart Innovation, Systems and Technologies), 445–457. doi:10.1007/978-81-322-2229-3_38

Svanæs, D. (2013). "Interaction design for and with the lived body: Some implications of merleau-ponty's phenomenology," in *ACM transactions on computer-human interaction (TOCHI)* (New York, NY, USA: ACM), 1–30.

Tan, H., Wang, D., and Sabanovic, S. (2018). "Projecting life onto robots: The effects of cultural factors and design type on multi-level evaluations of robot anthropomorphism," in 2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN) (Nanjing, China: IEEE), 129–136. doi:10.1109/ROMAN.2018.8525584

Tung, F.-W. (2016). Child perception of humanoid robot appearance and behavior. *Int. J. Human–Computer Interact.* 32, 493–502. doi:10.1080/10447318.2016.1172808

Unbehaun, D., Aal, K., Carros, F., Wieching, R., and Wulf, V. (2019). "Creative and cognitive activities in social assistive robots and older adults: Results from an exploratory field study with pepper," in Proceedings of the 17th European conference on computer-supported cooperative work-demos and posters (Salzburg, Austria: European Society for Socially Embedded Technologies).

Van Breemen, A. J. N. (2004). "Bringing robots to life: Applying principles of animation to robots," in Proceedings of shapping human-robot interaction workshop held at CHI (Vienna, Austria: Citeseer), 143–144.

Vignolo, A., Noceti, N., Rea, F., Sciutti, A., Odone, F., Sandini, G., et al. (2017). Detecting biological motion for human–robot interaction: A link between perception and action. *Front. Robot. AI* 4. doi:10.3389/frobt.2017.00014

WHO (2018). *Addictive behaviours*. New York, USA: Gaming disorder.

Won, A. S., Bailenson, J. N., Stathatos, S. C., and Dai, W. (2014). Automatically detected nonverbal behavior predicts creativity in collaborating dyads. *J. Nonverbal Behav.* 38, 389–408. doi:10.1007/s10919-014-0186-0

Wu, Y.-H., Fassert, C., and Rigaud, A.-S. (2012). Designing robots for the elderly: Appearance issue and beyond. *Archives gerontology geriatrics* 54, 121–126. doi:10.1016/j.archger.2011.02.003