

## APPLIED SCIENCES AND ENGINEERING

# Curriculum is more influential than haptic feedback when learning object manipulation

Pegah Ojaghi<sup>1†</sup>, Romina Mir<sup>2†</sup>, Ali Marjaninejad<sup>2,3</sup>, Andrew Erwin<sup>2,4,5</sup>, Michael Wehner<sup>6</sup>, Francisco J. Valero-Cuevas<sup>2,3,4,7,8\*</sup>

Dexterous manipulation remains an aspirational goal for autonomous robotic systems, particularly when learning to lift and rotate objects against gravity with intermittent finger contacts. We use model-free reinforcement learning to compare the effect of curriculum (i.e., combinations of lift and rotation tasks) and haptic information (i.e., no-tactile versus 3D-force) on learning with a simulated three-finger robotic hand. In addition, a novel curriculum-based learning rate scheduler accelerates convergence. We demonstrate that the choice of curriculum biases the progression of learning for dexterous manipulation across objects with different weights, sizes, and shapes—underscoring the robustness of our learning approach. Unexpectedly, learning is achieved even in the absence of haptic information. This challenges conventional thinking about task “complexity” and the necessity of haptic information for dexterous manipulation for this task. This work invites the analogy of curriculum learning as a malleable developmental process from a pluripotent state driven by the nature of the learning experience.

## INTRODUCTION

Dexterous manipulation is a triumph of biology (1–8). However, the autonomous learning of such behavior continues to remain out of reach for robots (4, 9–12). Robots have excelled at grasping, reaching for and statically coupling an object to the hand by applying forces with the fingertips, fingers, and palm (4, 6, 13, 14), for decades (15–23). But grasp is not dexterous manipulation (4). Dexterous in-hand manipulation, dynamically holding and reorienting an object with the fingertips (4, 18, 24, 25), is critical for interaction with, and use of, objects in unstructured human environments.

To achieve this kind of manipulation with multifingered robotic hands, the robotics community has developed sophisticated control theoretical approaches (4, 7, 18, 26–32); henceforth, we use the shorthand manipulation to mean dexterous in-hand manipulation. These control theoretical approaches, however, tend to require accurate models and state estimation, have narrow stability margins, and have difficulty compensating for friction, interpreting intermittent/deformable contact, and coordinating between multiple fingers. As an alternative approach, biorobotic, neuromechanics, and artificial intelligence communities have introduced a variety of bio-inspired and data-driven machine learning approaches in simulation and hardware (4, 11, 14, 33, 34).

One particularly promising approach is the subfield of reinforcement learning (RL), which has provided several successful examples

(12, 23, 35–39). RL empowers robots to iteratively enhance their manipulation skills through trial and error (without a need for an accurate model of the task or the environment), resulting in gradual improvements within complex environments. However, manipulation RL studies to date are usually highly computationally intensive—and have relied on vision—which limits their applicability (27, 34, 40–49). Last, most studies have been limited to the upward-facing hand configuration, relying on the palm as a resting platform for the object being manipulated which makes it an inherently more stable task to handle than a down-facing hand configuration (9). Adding the downward-facing hand configuration broadens the scope of solutions, delivering valuable insight to the robot manipulation community (9, 12, 50). However, it introduces additional challenges as this orientation requires the hand to counteract gravity at all times (51), and errors can lead to instabilities and failure by dropping the object. Here, we use an RL approach based on the proximal policy optimization (PPO) algorithm (52) to autonomously learn manipulation with a downward-facing hand without direct vision. We find that the choice of curriculum biases learning manipulation toward one or another combination of skills (i.e., lifting the ball and/or rotating it) more profoundly than the availability of tactile information.

Unexpectedly, the absence of tactile information did not necessarily prevent or substantially degrade learning relative to the influence of curriculum. These results reveal fundamental and previously underappreciated aspects of curricula as a powerful tool for autonomous learning of multiobjective tasks. For example, curricula commencing with both lift and rotation exhibit initial superior performance compared to those building up from simpler blocks, such as focusing solely on lift or rotation. Focusing on a single skill thereafter, however, can be additionally beneficial. Beyond assessing the impact of curricula on autonomous manipulation, our study yielded the significant revelation that, contrary to long-held notions, the absence of tactile information (and direct vision) does not inherently impede or degrade the learning process. There seems to be a functional interaction with a curriculum where available sensing capabilities bias the learning process toward combinations of dexterous manipulation skills that can leverage the available tactile information.

<sup>1</sup>Computer Science and Engineering Department, University of California Santa Cruz, Santa Cruz, CA, USA. <sup>2</sup>Alfred E. Mann Department of Biomedical Engineering, University of Southern California, Los Angeles, CA, USA. <sup>3</sup>Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA. <sup>4</sup>Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA, USA. <sup>5</sup>Mechanical and Materials Engineering Department, University of Cincinnati, Cincinnati, OH, USA. <sup>6</sup>Mechanical Engineering Department, University of Wisconsin-Madison, Madison, WI, USA. <sup>7</sup>Department of Aerospace and Mechanical Engineering, University of Southern California, Los Angeles, CA, USA. <sup>8</sup>Thomas Lord Department of Computer Science, University of Southern California, Los Angeles, CA, USA.

\*Corresponding author. Email: valero@usc.edu

†These authors contributed equally to this work.

**RESULTS**

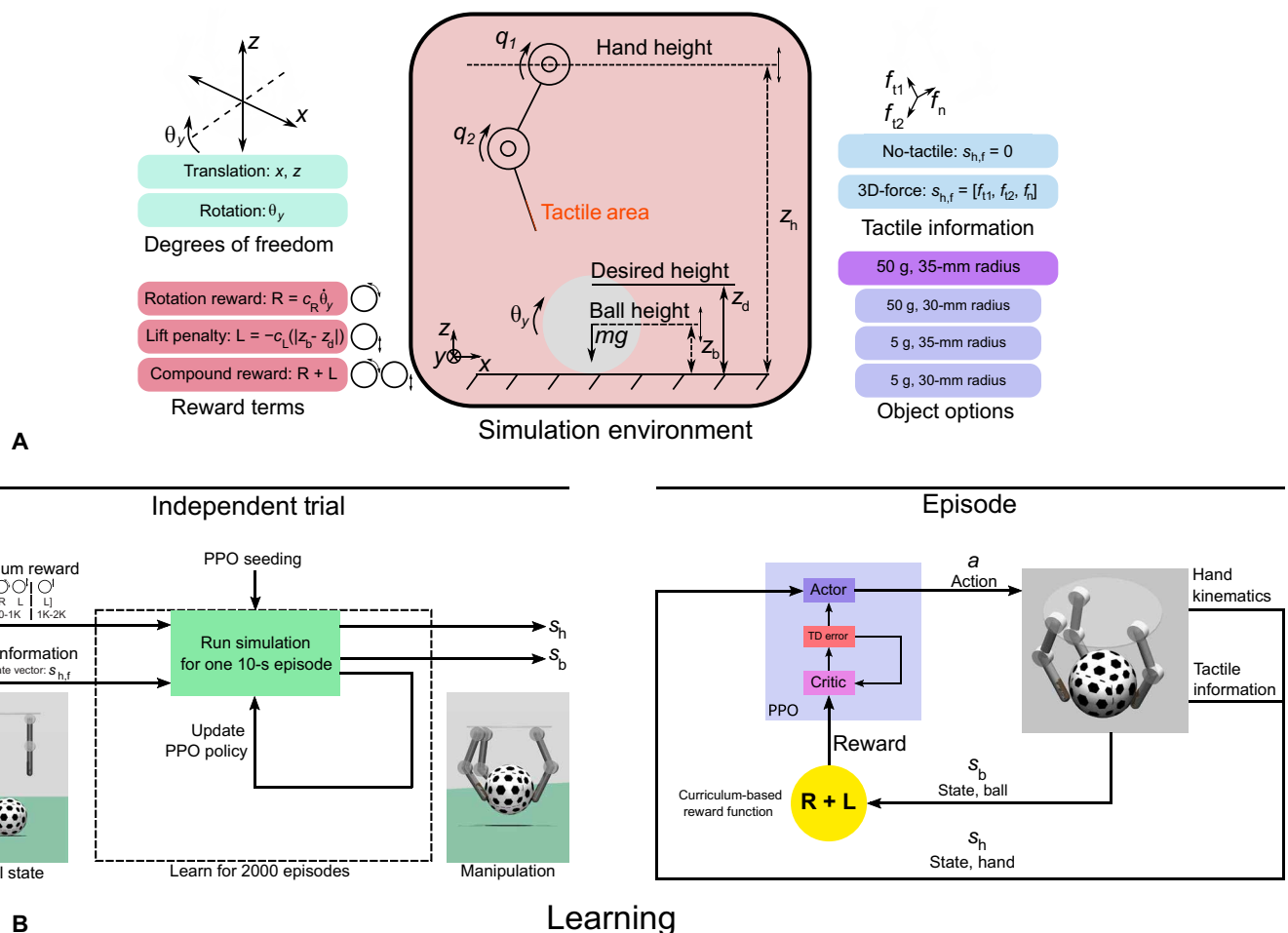
The goal of this project was to use curriculum-based RL with a simulated three-finger robotic hand to learn in-hand manipulation of an object against gravity in a data-efficient way—even while not using visual information. We demonstrate how the choice of curriculum is more influential than tactile information when learning to lift and rotate a ball (weighing 50 g with 35-mm radius) with a three-finger robotic hand in simulation (Fig. 1). To do so, we systematically explored two tactile conditions: no-tactile (no force perception at all at the fingertip) versus three-dimensional (3D)-force (a 3D-force vector in the direction of force at the fingertip) during five distinct curricula (details in Methods).

We defined each curriculum as implementing a learning policy that rewards various combinations and sequences of lift (L) and rotation (R) of a ball, which can switch at the halfway point (Methods).

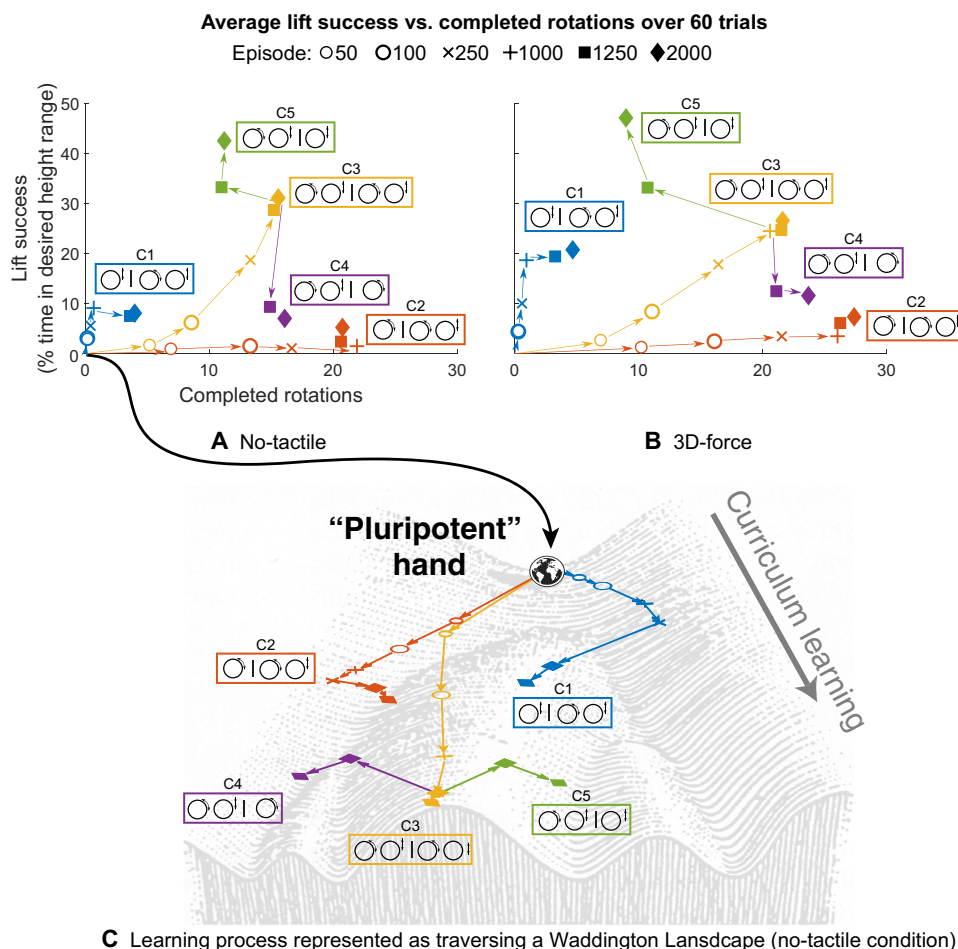
For example, curriculum 1 (i.e., C1) only rewards lift (L) in the first half of the trial, and both lift and rotation (L + R) in the second half are described as [L|L + R]. We find that the order of reward (curriculum) greatly affects the progression of learning and the final performance, 3D-force was not consistently better than no-tactile information, and a similar trend was observed across all configurations (see movie S1).

**Curriculum profoundly affects the progression of learning and final performance**

Each combination of curriculum and tactile information (Methods) leads to a distinct evolution of learning and final performance. This effect of curriculum affects both the progression of learning (path) and final performance (endpoint) and can be visualized as traversing a developmental process (as “Waddington Landscapes” in biology; Fig. 2; see Discussion).



**Fig. 1. Overview of simulation environment and learning.** A high-level overview of the simulation environment and learning approach to autonomous manipulation. See Methods for further details. **(A)** Simulation environment. A simulated three-finger robotic hand attempted to lift and rotate (i.e., dexterously manipulate) a ball with a weight of 50 g and 35-mm radius. The 3D movement of the ball was lightly constrained to the X-Z plane. Changes in the ball state affect the reward, which is a function of rotation, lift, and/or a combination of the two. We tested this approach with two different tactile information conditions (no-tactile and 3D-force) available at the fingertips and four balls of different weights and sizes. **(B)** Learning algorithm. Independent trial (left): For each of the five curricula, autonomous learning was evaluated over 60 independent trials (one trial shown). Each trial in a curriculum consisted of two learning phases lasting 1000 episodes for a total of 2000 episodes. The reward function changed at the end of the first learning phase (with the exception of curriculum 3; see Fig. 6). Episode (right): Each episode lasted 10 s and began de novo with the ball on the ground with the hand and fingertips suspended above it. The learning process was driven by PPO, utilizing an Actor-Critic architecture and Temporal Difference (TD) learning to dynamically update the agent’s actions (i.e., moving the fingers and hand) to maximize the curriculum’s reward.



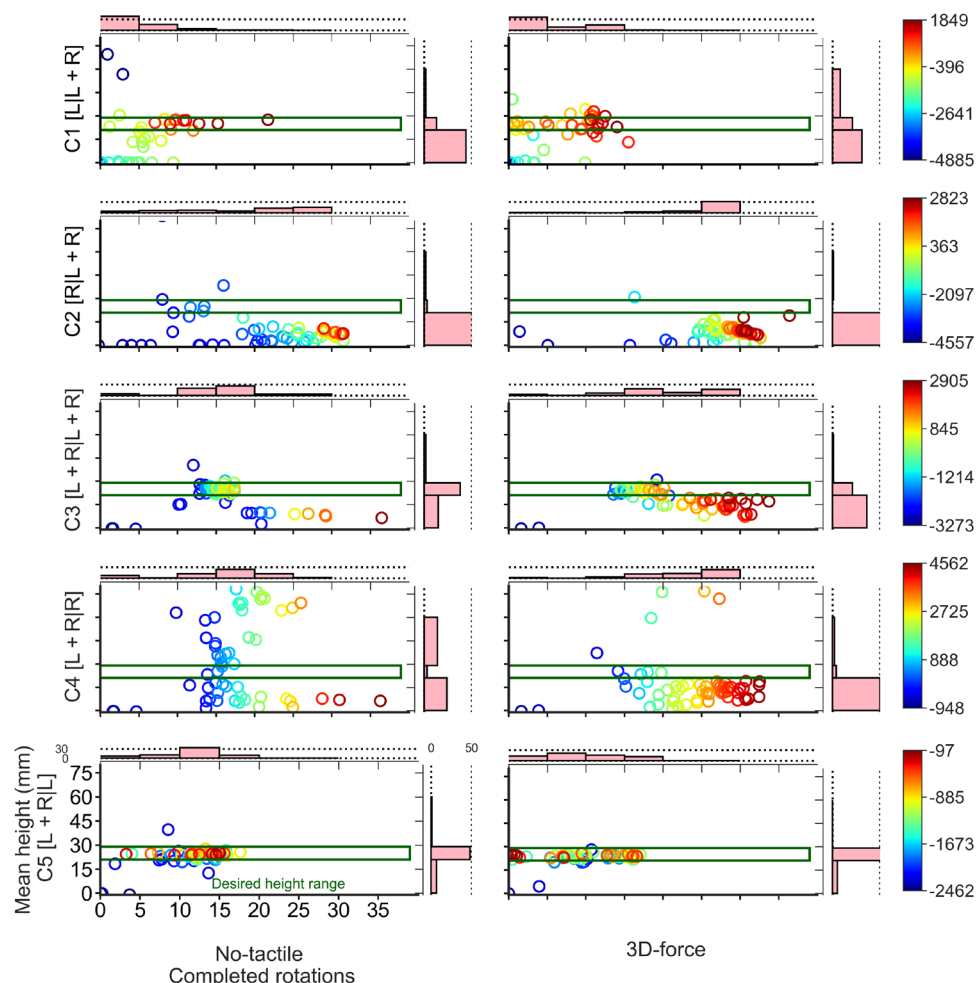
**Fig. 2. The evolution of learning highlights the dynamic functional interaction between curriculum and tactile information.** Manipulation performance during the last 10 s of each episode noted: The percent of the time the ball is within the desired height range versus number of complete rotations. Each point is the average of 60 independent trials. Arrows point in the direction of increasing episodes. Negative rotations were set to zero. Note that the choice of curriculum had a profound effect on learning for both tactile conditions [(A) no-tactile and (B) 3D-force]. Unexpectedly, learning happened even in the absence of tactile information, and manipulation performance was not always better with 3D-force information. (C) An analogy of learning as a developmental trajectory from a pluripotent state based on experience (curriculum). This effect of curriculum [and tactile information, cf. (A) versus (B)] affects both learning (path) and final performance (endpoint) and can be visualized as traversing a “Waddington Landscape” [adapted from (58)].

Curricula, as expected, diverge in their ability to lift and rotate the ball. They had the profound effect of biasing toward one or another combination of skills (L or R) and also adapting to the available sensory input, much like experience-dependent developmental paths from an initial pluripotent state (Fig. 2C). As we describe in detail in Discussion, we explicitly explored different initial rewards with similar final rewards (C1 [L|L + R] versus C2 [R|L + R] and vice versa (C4 [L + R|R] versus C5 [L + R|L])). In all cases, the system was able to respond to the change in reward (albeit with variable success). Note the evolution of skills for each curriculum tended to saturate quickly within the first 250 episodes of the first and second phases of learning. They tended to asymptote between the 250 and 1000 and between the 1250 and 2000 episodes, respectively. Nevertheless, the final endpoints for each curriculum differed substantially, showing that curricula are more than simply a means to learn multiobjective tasks, yet it can actually produce different learning paths and endpoints—which can be exploited by the user to achieve different capabilities with the same naïve system (Fig. 2).

Counterintuitively, starting with a multiobjective reward can be as effective, if not more effective, than starting with simpler rewards. For example, rewarding both lift and rotation during the first 1000 episodes (C3 [L + R|L + R], C4 [L + R|R], and C5 [L + R|L]) improves rotating the ball at the end of learning (episode 2000) better than when only rewarding rotation (C2 [R|L + R]) at the start.

**Tactile information is not necessary but can affect learning**

Most unexpectedly, the absence of tactile information did not preclude learning. Moreover, learning with no-tactile information was comparable to the 3D-force information (Fig. 2). The presence or absence of 3D-force information did, however, change the learning paths and endpoints of each curriculum (Figs. 2 and 3)—although the effect was not uniform. For example, 3D-force information did produce more lifting than no-tactile in C1 [L|L + R] at the end of learning. However, this was reversed in C3 [L + R|L + R]; and tactile information did not affect C4



**Fig. 3. Performance across all curricula and both tactile information conditions.** The joint distribution illustrates the performance during the final 10-s episode of each of the 60 trials [showcasing the mean ball height (millimeters) versus the number of completed rotations]. The color-coded cumulative reward for the last episode of each run (refer to Eq. 1) corresponds to different curricula. Note that the final manipulation performance is represented by those points inside the green box defining the desired ball height ( $25 \pm 4$  mm).

[L + R|R] or C5 [L + R|L] much (Fig. 2). This nuanced effect of tactile information at the end of learning is also seen in Fig. 3, and, on average, during learning in Fig. 4. This interaction was also seen while learning with different objects (see details in the “Generalizability” section in the Supplementary Materials and Fig. 5).

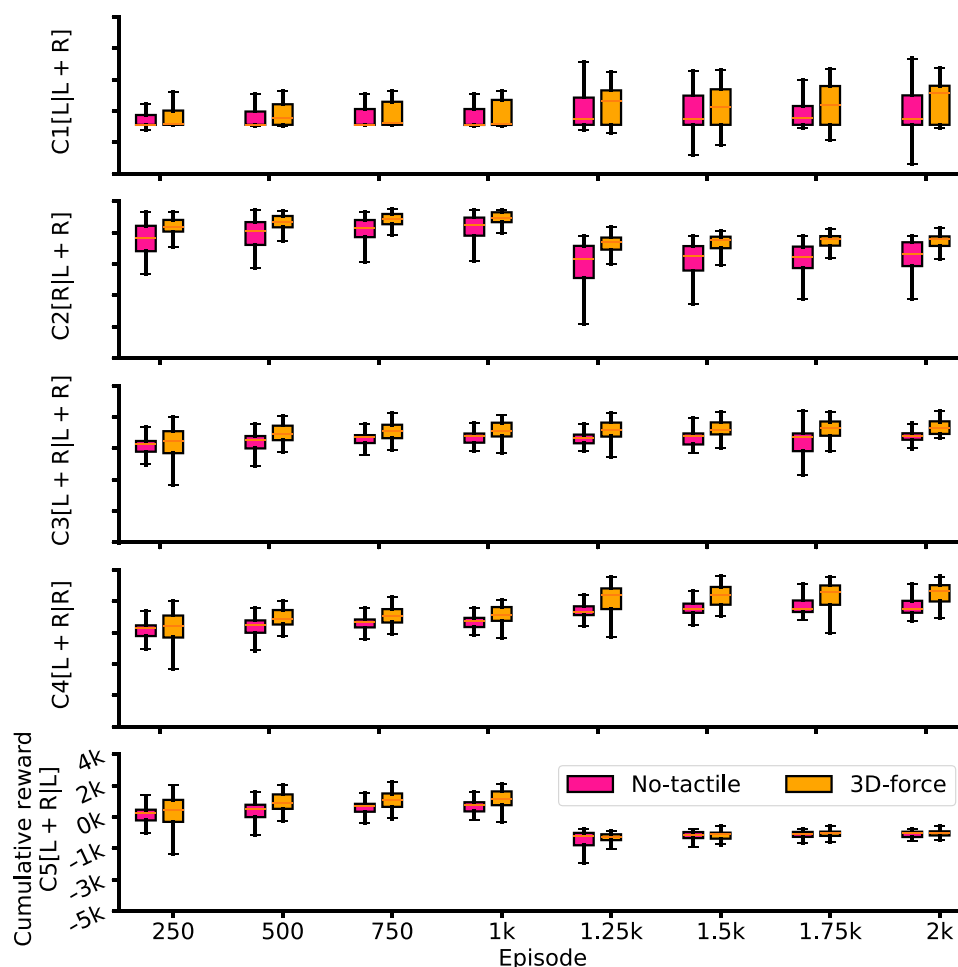
Further nuance of the effect of tactile information can be seen in the different paths of learning and in the response to switching of rewards between the first and second learning phases (i.e., after episode 1000). Note that C3 [L + R|L + R] rewards both skills during the entirety of both phases but tends to be most effective at lifting in the no-tactile information compared to 3D-force condition (Fig. 2). Nevertheless, when switching the reward to only lift C5 [L + R|L] or only rotation C4 [L + R|R] at the end of the first learning phase, the 3D-force case makes up for lost ground and has endpoints similar to those for the no-tactile information. This effect seems to be reversed for C1 [L|L + R] and C2 [R|L + R] where only lift or rotation was rewarded at first. In these cases, the 3D-force information produced greater lift and rotation during both learning phases.

## DISCUSSION

### What did we learn about learning to manipulate?

#### *Using a simulated three-finger robotic hand, we provide proof of principle that it is possible to learn the hard problem of dynamic dexterous manipulation*

Putting our work in context is critical and best done by pointing to its place in the updated taxonomy of hand function put forth by MacKenzie, Iberall, Brand, Cutkosky, Dollar, and others (2, 18, 53–57). In particular, we have addressed the problem of dynamic manipulation with three fingers, while the ball is at risk of being dropped at any moment (see the “Comparison to state of the art” section). This definition emphasizes that “grasp” and “pick-and-place manipulation” are conceptually and mechanically distinct from “dynamic manipulation” as addressed here, although they are at times used interchangeably in the literature (3). Such dynamic manipulation is an enviable ability that is also difficult for biology to achieve as it develops in humans late in childhood, degrades in healthy aging, and is quickly lost in even mild/initial forms of neurological conditions such as peripheral neuropathies, stroke, and Parkinson’s disease (3, 56). In our work, the fingertips induce



**Fig. 4. Cumulative reward across all curricula and tactile information conditions.** Boxplots, with median, across tactile conditions for 60 runs, every 250 episodes. Note that learning tends to saturate early.

dynamic translation and rotation of the ball while making and breaking contact. As such, the hand function we achieved merits the description of dynamic dexterous manipulation.

#### **Curriculum learning can be seen as a developmental process from a pluripotent state**

We use the analogy of the Waddington Landscape (Fig. 2C) for curriculum learning of manipulation (Fig. 6) because of its similarity to epigenetic transformation from a pluripotent state in biological development (57–59). Curriculum learning produces a developmental trajectory from a naïve (i.e., pluripotent) state based on resources (tactile information) and experience (curriculum) (Fig. 2, A and B). Each curriculum affects both the progression of the learning (path) and its final performance (endpoint) and can thus be thought of as traversing a Waddington Landscape.

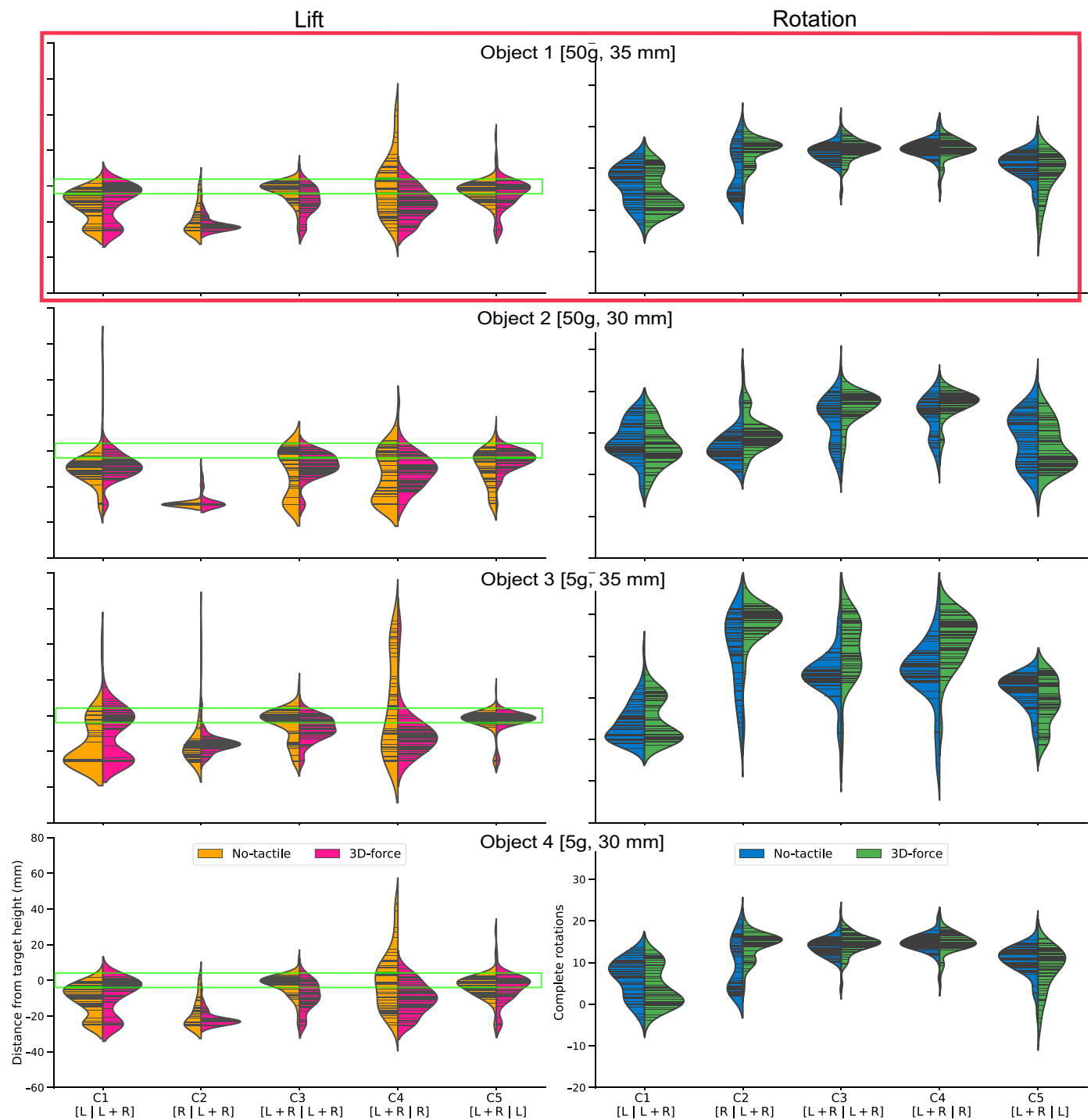
The evolution of skills for each curriculum was (unlike cell differentiation) not strictly irreversible, yet it remained adaptable. Specifically, the change of reward after the first learning phase did not preclude the system from emphasizing the improvement of the new skill. This is visually represented by 90° shifts in the paths (see C1 [L|L + R] and C2 [R|R + R] in Fig. 2). In some cases, the response to a switch in reward even reversed a learned skill for the first 250 episodes in the second phase of learning, and only then increased the

new skill (see C4 [L + R|R] and C5 [L + R|L] in Fig. 2). In one case, C3 [L + R|L + R], there was no change in reward after the end of the first learning phase, and the system was saturated already. In others, the system did respond like an “irreversible” system that learned little of the new skill, of at all, when the reward function was switched (e.g., C2 [R|R + R] in the 3D-tactile case in Fig. 2). See the next discussion section.

#### **The role of sensory information**











##### **Manipulation can be achieved without tactile information or vision**

Tactile information has long been thought as necessary for human—and by extension robotic—manipulation (4, 60). This idea was reinforced by the work of Johansson and Westling (61, 62) demonstrating that numbing the fingerpads with anesthetic temporarily impairs fine manipulation. However, the ability of individuals to still manipulate objects effectively despite impaired sensation (such as when wearing gloves in cold weather or with soapy hands) challenges the longstanding belief that tactile input is indispensable (61, 63–65). Our results in Fig. 2 provide a counter-example to this longstanding notion. We found that our system was able to learn even in the absence of tactile information (the no-tactile information



Downloaded from https://www.science.org on April 03, 2025

**Fig. 5.** Violin plots show the distribution of lift and rotation at the end of learning (i.e., the last 10 s of the 2000th episode) for all 60 trials. Final performance for lift (left) and rotation (right) for both tactile conditions for the ball shown in Fig. 2 and three others of different weights and radii shown. The top row corresponds to the reference ball described in the main results. The other balls are described in the Supplementary Materials. Lift is described as a distance from the desired height (the green box shows the distance from the desired height range  $\pm 4$  mm) and rotation as the number of completed rotations for both tactile conditions, no-tactile and 3D-force.

	Reward during first half	Reward during second half	Coefficients of equation [first — second] halves
Curriculum 1			$c_R = 0, c_L = 0.49$ — $c_R = 0.51, c_L = 0.49$ [L — L + R]
Curriculum 2			$c_R = 0.51, c_L = 0$ — $c_R = 0.51, c_L = 0.49$ [R — L + R]
Curriculum 3			$c_R = 0.51, c_L = 0.49$ — $c_R = 0.51, c_L = 0.49$ [L + R — L + R]
Curriculum 4			$c_R = 0.51, c_L = 0.49$ — $c_R = 0.51, c_L = 0$ [L + R — R]
Curriculum 5			$c_R = 0.51, c_L = 0.49$ — $c_R = 0, c_L = 0.49$ [L + R — L]

**Fig. 6. Curriculum definitions with combinations of two different subtasks of lift and rotation.** We used five curricula that rewarded different combinations of lifting and rotation during each half of the independent trials. These changes in the coefficients of the reward function define a progression of goals (i.e., curriculum learning) over the two halves of each run.

in Figs. 2A and 3). Having 3D-force not always produced better performance [cf. C3 [L + R|L + R] in Fig. 2 (A and B)].

How is it possible to learn to manipulate without vision or tactile information? The answer, we believe, comes from the nature of reinforcement itself. As described in Fig. 1B, PPO—as an RL algorithm—conditions its actions (next-step finger joint angles, angular velocities, palm position, and velocity) based on the system state, ultimately optimizing for increased reward. In the no-tactile case, the hand's state comprises finger joint angles, angular velocities, palm position, and palm velocity—which seem to suffice to learn the task. Therefore, lift and rotation of the ball was a product of guided hand kinematics that properly affect ball dynamics to increase the reward in the no-tactile case, such as in our previous work to learn locomotor movements without the need to sense the ground (38).

As such, a main contribution of our work is to provide an existence proof that an agent using RL is able to learn a sophisticated manipulation behavior even in the absence of tactile information. Note that direct vision was not necessary either, as in other prior work (38). Our important result about dynamic manipulation provides impetus to revise our thinking about, and use of, tactile and visual information to allow freer thinking for engineers (and bio-roboticists) creating the next generation of dexterous hands and robots.

**The presence or absence of tactile information did, however, alter the progression of learning**

Information presented in Fig. 2 (A and B) (and fig. S1) indicates that while the sensory conditions did influence the learning process across curricula, their overall learning trajectories and endpoint performance were similar. The effect of tactile information was not systematic. The 3D-force cases were not consistently or necessarily better than the no-tactile cases or vice versa. Thus, curriculum is a dominant factor compared to tactile information.

From the computational perspective, one could have expected that when learning with a fixed number of episodes, 3D-force sensor would perform systematically worse because of the computational demands associated with extending the length of the hand state vector  $s_h$  by nine elements (three forces per finger) for the same PPO

algorithm architecture which now has to tune more weights (Fig. 1). However, 3D-force sensor cases at times outperformed the no-tactile cases [e.g., C1 [L|L + R] in Fig. 1 (A and B)], which strongly suggests that our comparisons across curricula and tactile conditions are not the result of an imbalance in computational demands for a fixed number of learning episodes (1000 per learning phase for a total of 2000). This is additionally supported by the fact that 3D-force cases also saturated their learning by the 250th episode (like the no-tactile cases).

Also, it is important to note that this study does not undermine the effectiveness of tactile information in many everyday tasks. It merely provides a proof of principle that it is possible to learn a specific task (i.e., the manipulation task of interest in this paper) without using tactile information; and with performance comparable to when tactile information is available. It is clear that many tasks exist for which sensory signals would either be crucial to perform, or would greatly enhance, either the learning speed for the task, the final performance, error correction, and/or their robustness and repeatability. These are beyond the scope of our work.

**What did we learn about learning?**

**Our system exhibits some important features of lifelong learning**

As defined in (11), our system shows transfer and adaptation because it reuses knowledge to improve performance and rapidly adapts to previously unseen skills as in C1 [L|L + R], C2 [R|L + R], C4 [L + R|R], and C5 [L + R|L] (in Figs. 2 to 4). Similarly, our system did not suffer from catastrophic forgetting as it was able to retain varying amounts of previously learned knowledge on a case-by-case basis (Fig. 2 and fig. S1). For example, C4 [L + R|R] and C5 [L + R|L] did not entirely forget to lift or rotate when they were no longer rewarded, respectively.

**Curriculum learning does not necessarily have to advance gradually from single-objective to multiobjective rewards**

In many applications such as locomotion, investigators have found that curriculum learning is indispensable to advance gradually

from single-objective (i.e., “simpler”) to multiobjective (i.e., “more complex”) rewards (66). This has led to curriculum learning becoming the standard approach in the field. From the traditional definitions of Vanilla or Progressive curriculum learning (67, 68), one might assume that first learning to lift the ball (a form of grasp) is “easier” than rotating it, which involves a dynamic behavior (4, 18) and a curriculum strategy in which rotation is learned only after lift is going to be a significantly more successful one. However, rewarding lift and rotation from the start does not hinder learning, as demonstrated by C3 [L + R|L + R]. It allowed transfer and adaptation for C4 [L + R|R] and C5 [L + R|L] to subsequently refine the single skill rewarded during the second phase of learning—albeit at the expense of some reduction of the nonrewarded skill. However, it is noteworthy that curricula that rewarded only one skill from the start (C1 [L|L + R] and C2 [R|L + R]) were not able to learn the second skill as efficiently during the second learning phase (rotation and lift, respectively).

Another aspect of lifelong learning involves the saturation of capacity causing learning to slow down (69, 70). Capacity saturation arises due to the fixed representational capacity of parametric models, including the PPO algorithm (70). We see this in our implementation of PPO—which increasingly fails to absorb additional knowledge from successive episodes. This is most evident in C3 [L + R|L + R] for the entire second phase of learning, as shown in Fig. 2. A learning model with more free parameters would theoretically be able to absorb additional knowledge from successive episodes.

#### **The curriculum-based learning rate scheduler enhances the efficiency of learning which accelerates convergence to higher reward**

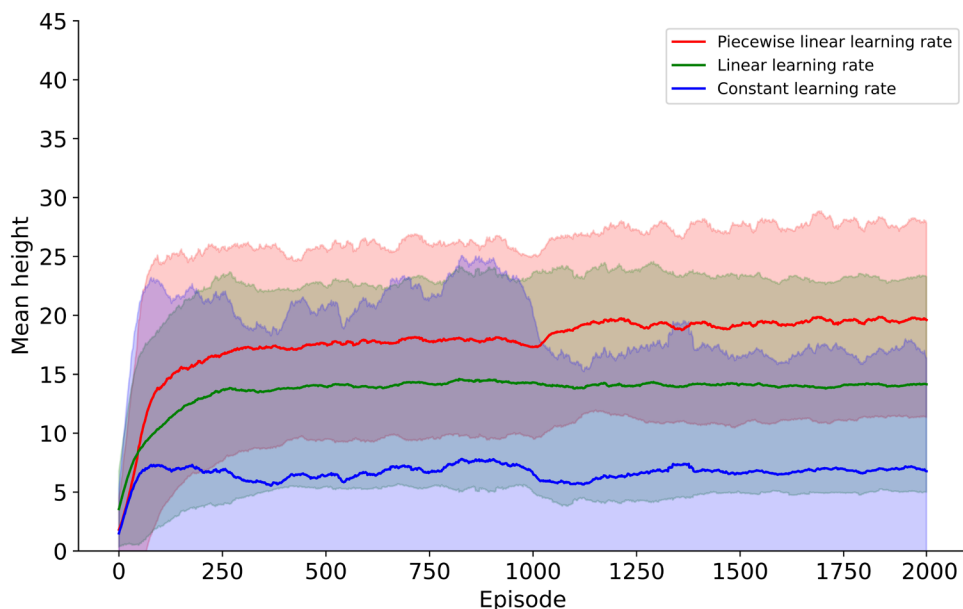
We sought to align the implementation of learning rates in PPO with the nature of curriculum learning. To do this, we defined our curriculum-based learning rate scheduler to adjust the linearly decaying learning rate when the reward changed (Fig. 7).

We find this improved learning and allowed a more fair comparison across curricula as it reduced heuristic tuning efforts. This curriculum-based learning rate scheduler offers an effective approach tailored to curriculum learning for autonomous systems by modifying the learning rate only when changing task complexities and rewards. Empowering curriculum learning to adapt learning rates in a way compatible with changing rewards enables autonomous systems to learn complex and dynamic environments more systematically, autonomously, and effectively. Thus, integrating curriculum-based learning and reward scheduling into a “curriculum-based learning rate scheduler” for autonomous systems is vital to enhance their learning capabilities and performance in manipulation tasks.

#### **Last, we demonstrate our results generalize to objects of different weights, shapes, textures, and sizes**

As shown in figs. S1 to S7, our results were consistent across four balls we studied (i.e., of two weights, 50 and 5 g, and two sizes, 35 and 30 mm in radius; Fig. 1A). There were minor differences across the endpoint performance for each object (note that the difference is the scales of the axis). But the learning paths for each curriculum and the effect of switching the reward remained consistent (fig. S1). This can also be seen in the detailed depiction of the distribution of rewards as learning progressed (Fig. 4). We further extended our results to manipulate objects with different textures and shapes. That is, we tested a softer ball (compared to the more rigid surface texture of the reference ball) and a cube. Details of these additional analyses can be found in the Supplementary Materials (table S5 and figs. S8 to S11).

The results for the cube are especially intriguing, as they reveal more distinct differences between training with and without tactile information. That is best seen in lower cumulative rewards across the entire training period for the no-tactile information than 3D-force for the cube (fig. S9) compared to the balls of similar weight



**Fig. 7. Effect of PPO curriculum-based learning rate scheduler by comparison of mean height.** Data presented for mean height in C5 [L + R|L] throughout the whole learning period. The desired height for all cases is 25 mm. Solid lines represent the mean across all 60 trials for the specified learning rate methods. Shaded areas represent  $\pm 1$ SD. The red solid line follows the PPO implementation per Eq. 2.



(Fig. 4 and fig. S3). Moreover, these same plots show greater dispersion in cumulative reward for the cube for both tactile conditions. This finding suggests that tactile information may become increasingly significant for both the level and consistency of performance as the complexity of the manipulated shape grows. Together, our work specific to dexterous manipulation with a three-finger robotic hand shows that our findings about curriculum learning and tactile conditions are robust to objects of different object sizes, weights, shapes, and textures.

### Comparison to the state of the art

It is critical to note that, as we have stated in the past (4), grasp or pick-and-place tasks are not dexterous manipulation in the rigorous sense of grasp taxonomies. Even reorienting a cube resting on an upward-facing palm (9, 16, 50, 71–74) is not prehensile manipulation. Moreover, prior work has relied heavily on extensive visual input for object reorientation on an upright palm (12, 75), with few exceptions such as Sievers *et al.*'s (76) demonstration of learning while slowly adding gravitational force. Notably, Chen *et al.* (12) recently demonstrated success in reorienting an irregular object with a downward-facing hand, although this approach remains dependent on vision. In addition, these approaches find it challenging to reorient unmarked symmetric objects.

Thus, we demonstrate a dynamic manipulation task against full gravity from the start, using curriculum learning where direct vision is not needed. We only required information about the height of the ball and its orientation, which can in practice be obtained by sensors other than vision that are sensitive to occlusions by the object and fingers. We used a novel curriculum-based learning rate scheduler for PPO, which significantly enhances the success performance across all scenarios. We now discuss how our approach to manipulation compares and contrasts with other studies in robotics and RL. The state of the art of autonomous learning for in-hand manipulation is limited. Although important advances have been made using computationally intensive approaches in simulation and hardware (12, 27, 34, 41–43), these tend to be impractical for autonomous learning at the edge.

Augmenting RL for manipulation with imitation learning has shown some successes (12, 35–37), but collecting task-specific expert demonstrations from humans is often limited to specific objects or tasks, might not always be practical, requires specialized equipment, and can be time-consuming.

In contrast, we used a model-free data-driven approach because precise prior knowledge of the system, objects, and the environment is not always available, especially in unstructured environments. Although some other studies also use model-free RL methods for rotating objects with simulated fingers or a robotic hand (9, 77–79), we have overcome some of their drawbacks. In (9, 78), the orientation of an object was controlled while resting on an upward-facing palm. Thus, it did not have to be held against gravity as it was not at risk of being dropped at any time.

Some of these limitations were addressed by Chen *et al.* (79) in simulation by manipulating the object with the palm facing downward like we did, but gravity was introduced slowly as part of the curriculum. Moreover, to successfully manipulate the object, the authors found it important to “initialize the object in a stable configuration”—which we did not need. Similarly, Caggiano *et al.* (25) have shown that incorporating a variety of complex object shapes in

training can support limited generalization even when no visual or tactile information is available.

The way our work went beyond the state of the art, therefore, is by demonstrating a method with the ability to autonomously learn to manipulate an object against gravity while revealing the role of curriculum learning and tactile information in in-hand manipulation. Curriculum learning strategies have been successfully applied across various areas of machine learning, including recent implementations in robotics and manipulation tasks (80–82). Our findings now show that curriculum learning not only facilitates performance but a curriculum can itself influence the trajectory of the learning process.

In addition, the impact of learning rate scheduling on stochastic optimizer performance has been extensively investigated in recent research (83, 84). In our study, we specifically explore the effects of a constant and linear piecewise learning rate for PPO on the success of our architecture. After careful consideration, we have decided to proceed with the piecewise learning rate. This adaptive approach adjusts rates dynamically throughout training, speeding up the process with higher initial rates and ensuring stable convergence with lower rates later on.

Last, our work underlines the importance of curricula in manipulation and shows how the right choice of a curriculum can enhance performance and robustness across multiple tasks by exhibiting some important features of lifelong learning. In this study, we emphasize that dexterous manipulation is not a monolithic task but rather a collection of interrelated challenges that can benefit from diverse learning strategies and sensory inputs. Our findings suggest that the complexity of manipulation tasks requires tailored approaches, highlighting the necessity for adaptive learning paradigms that can benefit from a variety of learning curricula and types of tactile information.

### Limitations, opportunities, and future directions

While our work pushes the field of autonomous manipulation forward, it naturally has some limitations. First, our work is done in simulation for a three-finger robotic hand. However, as with many other studies looking to bridge the sim2real divide (38, 85), we used realistic physical constraints within a state-of-the-art physics engine (MuJoCo) that handles dynamic contacts and affects well. This is a foundation that will enable future hardware implementations. As to the geometry of our hand, it is common for useful robotic hands to have three fingers (27, 77), but hands with more fingers—and other manipulation tasks—remain to be explored. Curriculum learning has multiple varieties (68) that can adapt as learning progresses such as self-paced curriculum learning. In our case, our learning phases were of fixed duration, although the system tended to plateau. Thus, it could benefit from future implementations that adapt reward changes to minimize training time. Last, our manipulation tasks serve as a foundation for—but do not yet address—traditional use cases for activities of everyday life.

Our choices regarding PPO, curriculum design, hand and object structure, reward function, and other parameters were specifically tailored to address the scientific questions of interest within the scope of this paper and to establish a proof of concept. It is important to emphasize that our selections were not intended as universally applicable solutions. That is to say, to address a different need, a similar pipeline to this paper can be used, but different tasks, environments, or robotic structures might need to be

used. Also, different learning blocks (different than the RL technique or the adaptive curriculum-based learning rate scheduler function used in this paper) can be used that might serve best for another specific task or purpose. An extension of this work has begun to reveal that, when clustering the performance of trails within a given curriculum, we can detect the emergence of distinct “learning trends” (86). Thus, this work motivates and justifies several future directions for research to understand how curricula can interact within and across sensory conditions, objects, and learning trends to enhance robustness, generalization, and transfer learning.

## METHODS

In this section, we first describe the simulation environment and the task used in this study. Then, we elaborate on the learning policy that enabled autonomous manipulation.

### Simulation environment

The manipulation and machine learning communities have used the advanced physics simulation environment MuJoCo (87) for tasks involving autonomous manipulation. MuJoCo allowed us to implement RL algorithms on a robotic hand in a realistic environment that includes contact dynamics (including penetration) and gravitational acceleration (87, 88).

To demonstrate the adaptability and robustness of our proposed methodologies, we assessed the performance using four different objects. Our evaluations encompassed systematic exploration, considering two different weight combinations (50 and 5 g), as well as varying ball radii (35 and 30 mm). The work presented herein focuses on a ball of 50 g with a radius of 35 mm with the other configurations presented in the “Generalizability” section in the Supplementary Materials.

### Robotic hand design

We simulated a bio-inspired, three-fingered robotic hand with a palm and three identical servo-driven fingers: two adjacent fingers, analogous to the “index” and “middle” fingers, and one opposing them, analogous to the “thumb.” In contrast to our prior efforts (89), where we showcased the reach-to-manipulate capability with a downward-facing orientation using distinct curricula, we modified the hand design. Each finger consisted of two joints that could rotate about the  $y$  axis ( $q_1$  and  $q_2$  in Fig. 1A), similar to the flexion or extension seen in human fingers. The size of the palm and length of each “phalanx” was based on an average human hand (77, 90). An additional servo motor was included at the base of the hand, which provides translational motion in the vertical direction ( $z_h$ ).

### Fingertip tactile sensors

This work incorporated tactile information and RL, sometimes referred to as touch-augmented RL, as we covered the internal side (i.e., the “pads” of the fingertips) of the distal phalanx of each finger with tactile sensors. Contact regions were configured near the tips of each finger (tactile area; Fig. 1). Objects contacting the finger outside of these tactile areas (sites, in MuJoCo) are not perceived as tactile information by the learning algorithm (78, 87).

We used MuJoCo’s built-in features to record the 3D-force sensor on the fingertips of all three fingers. The 3D-force sensor sites provide a 3D array of three orthogonal forces (one normal and two tangential to the sensor site for each sensor) of scalar values representing

the 3D-force vector. Moreover, we have considered an additional case: no-tactile. In the no-tactile case, the state vector for the tactile information  $s_{h,f}$  is null (we do not consider the tactile information in learning). As shown in Fig. 1A, the possible contact tactile information at each fingertip is indicated by  $s_{h,f} = [f_{t,1}, f_{t,2}, f_n]$  and it depends on tactile sensing available at fingertips. See table 2 for more details on the tactile information.

### Task description

The robotic hand attempts to manipulate a 50 g, 35-mm radius ball, which starts each episode on the ground with the palm of the robotic hand at a height of 200 mm above the ground. The ball height  $z_b$  is defined in the center of the ball, and we specified a desired height for the ball  $z_d$  to be 25 mm above  $z_b$ . In other words, the desired height  $z_d$  is 60 mm above the ground. Through simulation constraints, the ball is limited to 2 translational degrees of freedom (DOFs; moving vertically  $z$  and horizontally  $x$ ) and 1 rotational DOF (rotation about the  $\theta_y$  direction; see Fig. 1). We included viscous damping in the translational and rotational DOFs of the ball to stabilize the simulation and prevent numerical instabilities for the simulation of the rigid fingers.

We further limited the ball’s movement in the  $x$  direction by adding stiffness to the ball. The details of the simulation parameters, including the robotic hand and the ball, are shown in table S3.

### Observation and action space

The system’s state vector includes the hand state vector ( $s_h$ ), consisting of 14 kinematic DOFs, along with the position and velocities of the hand’s palm ( $s_p$ ) (2 DOFs), and the position and velocity of the ball ( $s_b$ ) (6 DOFs). This 20D vector encapsulates joint angles ( $q_1$  to  $q_6$ ) and their derivatives, as well as the vertical height of the hand ( $z_h$ ) and its derivative, collectively describing the dynamic state of the system.

In addition, the ball state vector comprises vertical ( $z$ ) and horizontal ( $x$ ) translation and its rotation about the  $y$  axis ( $\theta_y$ ). No other translations or rotations are permitted (see table S4). The height of the hand,  $z_h$ , is actuated for the hand to reach for and manipulate the state of the ball ( $s_b$ ) by rotating ( $\theta_y$ ) and lifting ( $z_b$ ) it to a desired height ( $z_d$ ).

It is important to note that not all state variables are used in our RL policy (observation state). Specifically, the observation state omits details about the ball’s velocity and position, as explained in the following subsection. Furthermore, it is worth mentioning that the action space aligns with the observation state. When a 3D-force is introduced, the state of the system dynamically changes, augmenting the hand state with an additional nine data points.

### Autonomous learning approach

To autonomously learn in-hand manipulation of a ball against gravity through using tactile information, we used a model-free RL algorithm to learn the policy. We used PPO as our main algorithm as it presented a balance between the ease of implementation, sample complexity, and ease of adjustment, trying to update at each step to minimize the cost function while assuring that the new policies are not too far from last policies (52, 91). PPO has also been adopted as one of the default methods of OpenAI owing to its excellent performance (92, 93).

### Reward function

The reward engineering concept (a subset of RL) focuses on finding the most appropriate reward to maximize successful learning via reward shaping (94). Reward shaping involves carefully designing

reward functions that provide the agent with rewards for progress toward the goal.

In our work, we defined two goals, lift and rotation. Lift: Our desired height (center of the ball above the ground) is  $z_d = 25$  mm, shown in (Fig. 1). In our algorithm, the goal is reached when the agent supports the ball against gravity within a desired height range of [21, 29] mm, indicated with a green box in Fig. 3 (and figs. S2, S4, and S6). A range is used to accommodate height variation during rotation and manipulation tasks. For result metrics, we report the mean height of the ball and lift success as a percent time within an episode where the ball is in the desired height range. Rotation: For rotation, we calculated completed rotations as our performance measurement (as opposed to rotation reward or rotation in degrees). Since we care about manipulating the ball against gravity at the desired height range, we used a combination of primary (positive) reward and punishment (penalty proportional to the distance between the current height and the desired height as a negative reward) at every time step.

In our reward function, the angular velocity of the ball  $\dot{\theta}_y$  was the primary reward, and the absolute distance of the state from the reference state of having the ball at the fixed desired position ( $z_d = 25$  mm; Fig. 1) was the punishment. The reward function is described by

$$\text{Reward}_t = c_R \dot{\theta}_{y,t} - c_L |z_{h,t} - z_d| \quad (1)$$

where  $c_R = 0.51$  and  $c_L = 0.49$ .

We investigated learning strategies (here, curriculum) in which lift and rotation are both rewarded (L + R), strategies in which only lift is rewarded with rotation coefficient set to zero ( $c_R$ ) and strategies in which lift coefficient is set to zero ( $c_L$ ) and only rotation is rewarded (R). This is described in detail in the following section (see Fig. 6).

### Curriculum learning

A learning trial consisted of 2000 episodes, where each episode lasted 10 s. This resulted in a total simulated time of 5 hours and 33 min per trial. Each learning trial was split into two equal halves where the reward function changed between the two halves of the trial (Algorithm 1). We considered five distinct curricula that differed in the behavior (rotation and lift) rewarded in two halves of the trial. This is illustrated by a circle with a curved arrow (rotation) and a vertical arrow (lift) throughout the paper and pictured in the second column of Fig. 6). As shown in the last column, by changing  $c_R$  and  $c_L$  variables in Eq. 1, we update the reward function in two equal halves of the learning trial in each curriculum. The final column of the table gives the values of  $c_R$  and  $c_L$  used in Eq. 1, to update the reward function in the two halves of the learning trial in each curriculum.

Learning was evaluated over 60 trials for each of the five curricula. Each of these 60 trials was independent by varying the seed parameters of the PPO algorithm for our RL policy. This was repeated for the two tactile conditions (no-tactile and 3D-force). For each tactile condition, the initial seed for the random number generator was held constant across different curricula. For example, the first trial run seed was exactly the same for all curricula and both tactile conditions. Overall, we used independent trials to evaluate the effectiveness of our approach to autonomous manipulation.

### Adaptive curriculum-based learning rate scheduler

The impact of learning rate scheduling on the performance of stochastic optimizers has garnered considerable attention in recent research (83, 84). Traditional approaches, using a fixed and static learning rate throughout training, often struggle to attain optimal model performance. To address this limitation, diverse scheduling algorithms, such as polynomial decay, cosine decay, and warm-up, have been proposed, each tailored with distinctive forms (95).

#### Algorithm 1. Simulation with PPO.

- 1: **procedure** RUNSIMULATION
- 2:     **Initialize** simulation environment
- 3:     Set random seed for reproducibility
- 4:     Initialize policy and value networks
- 5:     Initialize optimizers
- 6:     Set hyper-parameters and simulation parameters
- 7:     Initialize replay buffer
- 8:     Set training iterations and mini-batches
- 9:     Set PPO-specific parameters
- 10:    **for**  $episode = 1$  **to** 2,000 **do**
- 11:       **Initialize** episode
- 12:       **for**  $t = 1$  **to** 1,000 **do**
- 13:          Sample and execute actions
- 14:          Store transition in the replay buffer
- 15:       **end for**
- 16:       Update the policy using PPO
- 17:    **end for**
- 18: **end procedure**

Current methodologies often rely on predefined principles, assuming specific scheduling rules based on empirical studies and domain knowledge. These approaches may not rigidly adhere to any existing rule to find the optimal learning rate scheduling for a particular problem.

In our exploration of PPO, we aim to transcend the constraints associated with a constant learning rate. Initially, we opted for a constant learning rate and a linear learning rate, commonly used approaches in RL algorithms (96). But implementing a constant learning rate in dynamic contexts, where sensitivity to the initial rate choice can result in unstable training or sub-optimal solutions, highlights the necessity for adaptive approaches. We proposed a new method to tackle challenges with fixed learning rates, especially in dynamic environments like our manipulation tasks. This is addressed by an adaptive curriculum-based learning rate scheduler, bringing multiple advantages. This adaptive strategy dynamically adjusts rates throughout training, expediting the learning process with higher initial rates and ensuring stable convergence through decrementing rates during later stages.

### Curriculum-based learning rate scheduler strategy

Instead of using a fixed or decreasing learning rate, our method embraces a curriculum-adaptive learning strategy. The adaptive curriculum-based learning rate scheduler (piecewise linear learning rate) strategy is described as follows

$$L_r = \begin{cases} \phi \cdot \left(1 - \frac{\text{sample number}}{1,000,000}\right), & \text{sample number} \leq 1,000,000 \\ \eta \cdot \left(1 - \frac{\text{sample number}}{2,000,000}\right), & \text{sample number} > 1,000,000 \end{cases} \quad (2)$$

The selection of optimal values for  $\phi$  and  $\eta$  was determined empirically and set to 1 and 0.98 respectively, ensuring adaptability across all five curricula. The curriculum-based learning rate scheduler ( $L_r$ ) is established and adjusted through trial and error to emphasize the significance of curriculum learning. These coefficients are then integrated into the PPO linear scheduler according to the following equation. Our curriculum dynamically changes at 1000 episodes (1,000,000 samples), compelling the learning rate to be piecewise linear to accommodate the variations in the dynamics of the reward and tasks. This adaptive strategy effectively responds to changes in the environment, contributing to the model's success.

To validate the effectiveness of our approach, we explore the impact of constant, linear, and adaptive curriculum-based learning rate scheduler (Piecewise Linear Learning Rate) in C5 [L + R|L] in Fig. 7, comparing mean height more than 2000 learning episodes. The piecewise linear learning rate was far closer to this target height than either the linear or constant learning rate. Thus, the piecewise linear learning rate was used as the curriculum-based learning rate scheduler throughout this work. Our results in different curricula consistently support the superior performance of PPO with well-designed scheduling mechanisms, surpassing those using a constant and linear learning rate in both convergence rate and final performance metrics (95, 97, 98). One key advantage is the reduced sensitivity to the initial rate choice, minimizing the risk of divergence. The piecewise linear learning rate promotes efficient exploration in the early stages and exploitation for optimal performance during convergence. Its curriculum-based adaptive nature contributes to faster convergence, effectively navigating both exploratory and exploitative learning phases. Moreover, the piecewise

linear schedule imparts robustness against variations in task difficulty or environmental changes, automatically adjusting to maintain training stability.

To evaluate the effectiveness of different learning scheduler methods in reducing convergence time, we conducted an analysis on the average number of episodes needed after switching the reward during the second phase of learning in C5 [L + R|L]. We compared three learning schedulers: constant learning rate, linear rate, and piecewise linear rate.

Our findings reveal that the average number of episodes for convergence in successful trials (defined as trials where the hand can maintain the ball within the target height range) after the reward switch varied significantly across the different schedulers. Specifically, when focusing only on the successful trials (not shown), we observed that it took 1000 episodes for convergence with a constant learning rate, 450 episodes with a linear rate, and only 250 episodes with a piecewise linear rate (see episodes 1250 in Fig. 7).

Figure 7 illustrates the performance of each scheduler in reaching the target height. The piecewise linear learning rate outperformed both the linear and constant rates by a substantial margin. In addition, it achieved a higher cumulative reward across all 60 trials, indicating its superior effectiveness in learning and adaptation. These results highlight the significant advantages of using a piecewise linear learning rate scheduler in enhancing convergence speed and overall performance in C5 [L + R|L] simulations.

In summary, our adaptive curriculum-based learning rate scheduler strategy in the PPO implementation aims to enhance training stability, expedite convergence, and improve adaptability in dynamic environments. This aligns with our goal of efficiently training the agent for effective in-hand manipulation and contributes to the exploration of learning rate scheduling strategies on a curriculum-based approach. The complete code for learning is available at the following <https://github.com/pojaghi/In-hand-manipulation>.

### Supplementary Materials

#### The PDF file includes:

Supplementary Methods  
Supplementary Results  
Figs. S1 to S11  
Tables S1 to S5  
Legend for movie S1

#### Other Supplementary Material for this manuscript includes the following:

Movie S1

### REFERENCES AND NOTES

1. R. N. Lemon, R. S. Johansson, G. Westling, Corticospinal control during reach, grasp, and precision lift in man. *J. Neurosci.* **15**, 6145–6156 (1995).
2. C. L. MacKenzie, T. Iberall, "The grasping hand" in *Advances in Psychology* (Elsevier, 1994), vol. 104.
3. F. J. Valero-Cuevas, Why the hand? *Adv. Exp. Med. Biol.* **629**, 553–557 (2009).
4. F. J. Valero-Cuevas, M. Santello, On neuromechanical approaches for the study of biological and robotic grasp and manipulation. *J. Neuroeng. Rehabil.* **14**, 1–20 (2017).
5. A. Billard, D. Kragic, Trends and challenges in robot manipulation. *Science* **364**, eaat8414 (2019).
6. V. Ortenzi, M. Controzzi, F. Cini, J. Leitner, M. Bianchi, M. A. Roa, P. Corke, Robotic manipulation and the role of the task in the metric of success. *Nat. Mach. Intell.* **1**, 340–346 (2019).
7. M. V. Liarokapis, A. M. Dollar, "Learning task-specific models for dexterous, in-hand manipulation with simple, adaptive robot hands" in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2016), pp. 2534–2541.

8. A. S. Sadun, J. Jalani, F. Jamil, "Grasping analysis for a 3-finger adaptive robot gripper" in *2016 2nd IEEE International Symposium on Robotics and Manufacturing Automation (ROMA)* (IEEE, 2016), pp. 1–6.
9. M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. Weng, W. Zaremba, Learning dexterous in-hand manipulation. *Int. J. Robot. Res.* **39**, 3–20 (2020).
10. A. Gupta, J. Yu, T. Z. Zhao, V. Kumar, A. Rovinsky, K. Xu, T. Devlin, S. Levine, "Reset-free reinforcement learning via multi-task learning: Learning dexterous manipulation behaviors without human intervention" in *2021 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2021), pp. 6664–6671.
11. D. Kudithipudi, M. Aguilar-Simon, J. Babb, M. Bazhenov, D. Blackiston, J. Bongard, A. P. Brna, S. C. Raja, N. Cheney, J. Clune, A. Daram, S. Fusi, P. Helfer, L. Kay, N. Ketz, Z. Kira, S. Kolouri, J. L. Krichmar, S. Kriegman, M. Levin, S. Madireddy, S. Manicka, A. Marjaninejad, B. M. Naughton, R. Miikkulainen, Z. Navratilova, T. Pandit, A. Parker, P. K. Pilly, S. Risi, T. J. Sejnowski, A. Soltoggio, N. Soares, A. S. Tolia, D. Urbina-Meléndez, F. J. Valero-Cuevas, G. M. van de Ven, J. T. Vogelstein, F. Wang, R. Weiss, A. Yanguas-Gil, X. Zou, H. Siegelmann, Biological underpinnings for lifelong learning machines. *Nat. Mach. Intell.* **4**, 196–210 (2022).
12. T. Chen, M. Tippur, S. Wu, V. Kumar, E. Adelson, P. Agrawal, Visual dexterity: In-hand reorientation of novel and complex object shapes. *Sci. Robot.* **8**, eadc9244 (2023).
13. S. Katyara, F. Ficuciello, D. G. Caldwell, B. Siciliano, F. Chen, Leveraging kernelized synergies on shared subspace for precision grasping and dexterous manipulation. *IEEE Trans. Cogn. Dev. Syst.* **15**, 2064–2076 (2021).
14. A. Bicchi, Hands for dexterous manipulation and robust grasping: A difficult road toward simplicity. *IEEE Trans. Robot. Autom.* **16**, 652–662 (2000).
15. A. Bicchi, V. Kumar, "Robotic grasping and contact: A review" in *IEEE International Conference on Robotics and Automation* (IEEE, 2000), vol. 1, pp. 348–353.
16. R. Deimel, O. Brock, A novel type of compliant and underactuated robotic hand for dexterous grasping. *Int. J. Robot. Res.* **35**, 161–185 (2016).
17. E. Brown, N. Rodenberg, J. Amend, A. Mozeika, E. Steltz, M. R. Zakin, H. Lipson, H. M. Jaeger, Universal robotic gripper based on the jamming of granular material. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18809–18814 (2010).
18. R. M. Murray, Z. Li, S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation* (CRC Press, 1994).
19. A. T. Miller, P. K. Allen, GraspIt! A versatile simulator for robotic grasping. *IEEE Robot. Autom. Mag.* **11**, 110–122 (2004).
20. A. M. Dollar, R. D. Howe, "The SDM hand as a prosthetic terminal device: A feasibility study" in *2007 IEEE 10th International Conference on Rehabilitation Robotics* (IEEE, 2007), pp. 978–983.
21. M. T. Mason, K. Y. Goldberg, R. H. Taylor, "Planning sequences of squeeze-grasps to orient and grasp polygonal objects" in *Seventh CISM-IFTOMM Symposium on Theory and Practice of Robots and Manipulators* (1988), pp. 88–127.
22. K. Zeissler, A robotic hand gets a grip. *Nat. Electron.* **5**, 18 (2022).
23. E. Triantafyllidis, F. Acero, Z. Liu, Z. Li, Hybrid hierarchical learning for solving complex sequential tasks using the robotic manipulation network roman. *Nat. Mach. Intell.* **5**, 991–1005 (2023).
24. M. R. Cutkosky, R. D. Howe, "Human grasp choice and robotic grasp analysis" in *Dextrous Robot Hands*, S. T. Venkataraman, T. Iberall, Eds. (Springer-Verlag, 1990), pp. 5–31.
25. V. Caggiano, G. Durandau, C. Wang, C. K. Tan, P. Schumacher, H. Wang, A. Chiappa, A. M. Vargas, A. Mathis, J. Park, J. Won, G. Park, B. Shin, M. Kim, S. Koo, Z. Yang, W. Dang, H. Cai, J. Song, S. Song, M. Sartori, V. Kumar, Myochallenge 2023: Towards human-level dexterity and agility (2024).
26. D. Guo, F. Sun, B. Fang, C. Yang, N. Xi, Robotic grasping using visual and tactile sensing. *Inform. Sci.* **417**, 274–286 (2017).
27. A. S. Morgan, K. Hang, B. Wen, K. Bekris, A. M. Dollar, Complex in-hand manipulation via compliance-enabled finger gaing and multi-modal planning. *IEEE Robot. Autom. Lett.* **7**, 4821–4828 (2022).
28. R. O. Ambrose, R. Hal Aldridge, S. Askew, R. R. Burridge, W. Bluethmann, M. Diftler, C. Lovchik, D. Magruder, F. Rehnmark, Robonaut: NASA's space humanoid. *IEEE Intell. Syst.* **15**, 57–63 (2000).
29. M. G. Catalano, G. Grioli, E. Farnioli, A. Serio, C. Piazza, A. Bicchi, Adaptive synergies for the design and control of the pisa/iit soft hand. *Int. J. Robot. Res.* **33**, 768–782 (2014).
30. M. Bridges, J. Beaty, F. Tenore, M. Para, M. Mashner, V. Aggarwal, S. Acharya, G. Singhal, N. Thakor, "Revolutionizing prosthetics 2009: Dexterous control of an upper-limb neuroprosthesis" in *Johns Hopkins APL Technical Digest* (Applied Physics Laboratory, 2010), vol. 28, pp. 210–211.
31. C. Castellini, P. Van Der Smagt, Surface EMG in advanced hand prosthetics. *Biol. Cybern.* **100**, 35–47 (2009).
32. M. Huber, R. A. Gruben, Robust finger gaits from closed-loop controllers. *IEEE/RSJ Int. Conf. Intell. Robots Syst.* **2**, 1578–1584 (2002).
33. F. J. Valero-Cuevas, H. Hoffmann, M. U. Kurse, J. J. Kutch, E. A. Theodorou, Computational models for neuromuscular function. *IEEE Rev. Biomed. Eng.* **2**, 110–135 (2009).
34. A. Rajeswaran, V. Kumar, A. Gupta, G. Vezzani, J. Schulman, E. Todorov, S. Levine, Learning complex dexterous manipulation with deep reinforcement learning and demonstrations. arXiv:1709.10087 [cs.LG] (2017).
35. R. Jeong, J. T. Springenberg, J. Kay, D. Zheng, Y. Zhou, A. Galashov, N. Heess, F. Nori, Learning dexterous manipulation from suboptimal experts. arXiv:2010.08587 [cs.RO] (2020).
36. H. Zhu, A. Gupta, A. Rajeswaran, S. Levine, V. Kumar, "Dexterous manipulation with deep reinforcement learning: Efficient, general, and low-cost" in *2019 International Conference on Robotics and Automation* (IEEE, 2019), pp. 3651–3657.
37. A. Gupta, C. Eppner, S. Levine, P. Abbeel, "Learning dexterous manipulation for a soft robotic hand from human demonstration" in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IEEE, 2016), pp. 3786–3793.
38. A. Marjaninejad, D. Urbina-Meléndez, B. A. Cohn, F. J. Valero-Cuevas, Autonomous functional movements in a tendon-driven limb via limited experience. *Nat. Mach. Intell.* **1**, 144–154 (2019).
39. V. Caggiano, S. Dasari, V. Kumar, "Myodex: A generalizable prior for dexterous manipulation" in *International Conference on Machine Learning* (PMLR, 2023), pp. 3327–3346.
40. S. Zhang, Y. Liu, J. Deng, X. Gao, J. Li, W. Wang, M. Xun, X. Ma, Q. Chang, J. Liu, W. Chen, J. Zhao, Piezo robotic hand for motion manipulation from micro to macro. *Nat. Commun.* **14**, 500 (2023).
41. V. Kumar, E. Todorov, "MuJoCo HAPTIX: A virtual reality system for hand manipulation" in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)* (IEEE, 2015), pp. 657–663.
42. N. Funk, C. Schaff, R. Madan, T. Yoneda, J. De Urain Jesus, J. Watson, E. K. Gordon, F. Widmaier, S. Bauer, S. S. Srinivasa, T. Bhattacharjee, M. R. Walter, J. Peters, Benchmarking structured policies and policy optimization for real-world dexterous object manipulation. *IEEE Robot. Autom. Lett.* **7**, 478–485 (2021).
43. S. Cruciani, B. Sundaralingam, K. Hang, V. Kumar, T. Hermans, D. Kragic, Benchmarking in-hand manipulation. *IEEE Robot. Autom. Lett.* **5**, 588–595 (2020).
44. E. Theodorou, E. Todorov, F. J. Valero-Cuevas, "Neuromuscular stochastic optimal control of a tendon driven index finger model" in *Proceedings of the 2011 American Control Conference* (IEEE, 2011), pp. 348–355.
45. H. Sun, K. J. Kuchenbecker, G. Martius, A soft thumb-sized vision-based sensor with accurate all-round force perception. *Nat. Mach. Intell.* **4**, 135–145 (2022).
46. M. Thor, P. Manoonpong, Versatile modular neural locomotion control with fast learning. *Nat. Mach. Intell.* **4**, 169–179 (2022).
47. J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, M. Hutter, Learning quadrupedal locomotion over challenging terrain. *Sci. Robot.* **5**, eabc5986 (2020).
48. M. Schilling, K. Konen, T. Korthals, "Modular deep reinforcement learning for emergent locomotion on a six-legged robot" in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechanics (BioRob)* (IEEE, 2020), pp. 946–953.
49. J. Clune, K. O. Stanley, R. T. Pennock, C. Ofria, On the performance of indirect encoding across the continuum of regularity. *IEEE Trans. Evol. Comput.* **15**, 346–367 (2011).
50. OpenAI, I. Akkaya, M. Andrychowicz, M. Chociej, M. Litwin, B. McGrew, A. Petron, A. Paino, M. Plappert, G. Powell, R. Ribas, J. Schneider, N. Tezak, J. Tworek, P. Welinder, L. Weng, Q. Yuan, W. Zaremba, L. Zhang, Solving rubik's cube with a robot hand. arXiv:1910.07113 [cs.LG] (2019).
51. W. Hu, B. Huang, W. W. Lee, S. Yang, Y. Zheng, Z. Li, Dexterous in-hand manipulation of slender cylindrical objects through deep reinforcement learning with tactile sensing. arXiv:2304.05141 [cs.RO] (2023).
52. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms. arXiv:1707.06347 [cs.LG] (2017).
53. I. M. Bullock, A. M. Dollar, "Classifying human manipulation behavior" in *2011 IEEE International Conference on Rehabilitation Robotics* (IEEE, 2011), pp. 1–6.
54. T. Feix, J. Romero, H. B. Schmedmayer, A. M. Dollar, D. Kragic, The grasp taxonomy of human grasp types. *IEEE Trans. Hum.-Mach. Syst.* **46**, 66–77 (2015).
55. M. R. Cutkosky, On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Trans. Robot. Autom.* **5**, 269–279 (1989).
56. P. W. Brand, "Clinical mechanics of the hand" in *Hand Rehabilitation in Occupational Therapy* (Routledge, 2012), pp. 183–184.
57. M. T. Duroz, "Assessment of hand functions" in *Hand function: A Practical Guide to Assessment* (Springer, 2014), pp. 41–51.
58. C. H. Waddington, Evolutionary adaptation. *Perspect. Biol. Med.* **2**, 379–401 (1959).
59. C. Guerrero-Bosagna, J. Lees, D. Núñez-León, J. F. Botelho, "Epigenetics, evolution and development of birds" in *Epigenetics, Development, Ecology and Evolution* (Springer, 2022), pp. 149–176.

60. N. Wettels, V. J. Santos, R. S. Johansson, G. E. Loeb, Biomimetic tactile sensor array. *Adv. Robot.* **22**, 829–849 (2008).
61. R. S. Johansson, G. Westling, Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects. *Exp. Brain Res.* **56**, 550–564 (1984).
62. R. S. Johansson, C. Häger, R. Riso, Somatosensory control of precision grip during unpredictable pulling loads. *Exp. Brain Res.* **89**, 192–203 (1992).
63. R. S. Johansson, K. J. Cole, Grasp stability during manipulative actions. *Can. J. Physiol. Pharmacol.* **72**, 511–524 (1994).
64. D. A. Nowak, J. Hermsdörfer, S. Glasauer, J. Philipp, L. Meyer, N. Mai, The effects of digital anaesthesia on predictive grip force adjustments during vertical movements of a grasped object. *Eur. J. Neurosci.* **14**, 756–762 (2001).
65. E. Pavlova, Å. Hedberg, E. Ponten, S. Gantelius, F. J. Valero-Cuevas, H. Forssberg, Activity in the brain network for dynamic manipulation of unstable objects is robust to acute tactile nerve block: An fmri study. *Brain Res.* **1620**, 98–106 (2015).
66. P. Ramdya, A. J. Ijspeert, The neuromechanics of animal locomotion: From biology to robotics and back. *Sci. Robot.* **8**, eadg0279 (2023).
67. Y. Bengio, J. Louradour, R. Collobert, J. Weston, “Curriculum learning” in *Proceedings of the 26th Annual International Conference on Machine Learning (PMLR, 2009)*, pp. 41–48.
68. P. Saviary, R. T. Ionescu, P. Rota, N. Sebe, Curriculum learning: A survey. *Int. J. Comput. Vis.* **130**, 1526–1565 (2022).
69. M. McCloskey, N. J. Cohen, Catastrophic interference in connectionist networks: The sequential learning problem. *Psychol. Learn. Motiv.* **24**, 109–165 (1989).
70. S. Sodhani, S. Chandar, Y. Bengio, Toward training recurrent neural networks for lifelong learning. *Neural Comput.* **32**, 1–35 (2020).
71. Z.-H. Yin, B. Huang, Y. Qin, Q. Chen, X. Wang, Rotating without seeing: Towards in-hand dexterity through touch. arXiv:2303.10880 [cs.RO] (2023).
72. H. Qi, B. Yi, S. Suresh, M. Lambeta, Y. Ma, R. Calandra, J. Malik, General in-hand object rotation with vision and touch. arXiv:2309.09979 [cs.RO] (2023).
73. I. Shenfeld, Z.-W. Hong, A. Tamar, P. Agrawal, “Tgrl: Teacher guided reinforcement learning algorithm for pomdps” in *Workshop on Reincarnating Reinforcement Learning at ICLR 2023 (ICLR, 2023)*.
74. L. Yang, B. Huang, Q. Li, Y.-Y. Tsai, W. W. Lee, C. Song, J. Pan, Tacgnn: Learning tactile-based in-hand manipulation with a blind robot using hierarchical graph neural network. *IEEE Robot. Autom. Lett.* **8**, 3605–3612 (2023).
75. M. Andrychowicz, M. Denil, S. Gomez, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, N. De Freitas, “Learning to learn by gradient descent by gradient descent” in *Advances in Neural Information Processing Systems 29 (NIPS, 2016)*.
76. L. Sievers, J. Pitz, B. Bäuml, “Learning purely tactile in-hand manipulation with a torque-controlled hand” in *2022 International Conference on Robotics and Automation (IEEE, 2022)*, pp. 2745–2751.
77. H. Van Hoof, T. Hermans, G. Neumann, J. Peters, “Learning robot in-hand manipulation with tactile features” in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids) (IEEE, 2015)*, pp. 121–127.
78. A. Melnik, L. Lach, M. Plappert, T. Korthals, R. Haschke, H. Ritter, “Tactile sensing and deep reinforcement learning for in-hand manipulation tasks” in *IROS Workshop on Autonomous Object Manipulation (IEEE, 2019)*.
79. T. Chen, J. Xu, P. Agrawal, “A system for general in-hand object re-orientation” in *Conference on Robot Learning (PMLR, 2022)*, pp. 297–307.
80. E. Sayar, G. Iacca, A. Knoll, Curriculum learning for robot manipulation tasks with sparse reward through environment shifts. *IEEE Access* **12**, 46626–46635 (2024).
81. S. Dasari, A. Gupta, V. Kumar, “Learning dexterous manipulation from exemplar object trajectories and pre-grasps” in *2023 IEEE International Conference on Robotics and Automation (ICRA) (2023)*, pp. 3889–3896.
82. Z. Wang, Y. Jia, L. Shi, H. Wang, H. Zhao, X. Li, J. Zhou, J. Ma, G. Zhou, Arm-constrained curriculum learning for loco-manipulation of the wheel-legged robot. arXiv:2403.16535 [cs.RO] (2024).
83. A. F. Cooper, Y. Lu, J. Forde, C. M. De Sa, Hyperparameter optimization is deceiving us, and how to stop it. *Adv. Neural Inf. Process. Syst.* **34**, 3081–3095 (2021).
84. Z. Xu, A. M. Dai, J. Kemp, L. Metz, Learning an adaptive learning rate schedule. arXiv:1909.09712 [cs.LG] (2019).
85. F. Ruppert, A. Badri-Spröwitz, Learning plastic matching of robot dynamics in closed-loop central pattern generators. *Nat. Mach. Intell.* **4**, 652–660 (2022).
86. R. Mir, P. Ojaghi, A. Erwin, A. Marjaninejad, M. Wehner, F. J. Valero-Cuevas, J. Francisco, “Curriculum learning influences the emergence of different learning trends” in *2024 IEEE-RAS (Robotics and Automation Society) International Conference on Humanoid Robots (IEEE, 2024)*, pp. 521–527.
87. E. Todorov, T. Erez, Y. Tassa, “MuJoCo: A physics engine for model-based control” in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IEEE, 2012)*, pp. 5026–5033.
88. V. Kumar, Y. Tassa, T. Erez, E. Todorov, “Real-time behaviour synthesis for dynamic hand-manipulation” in *2014 IEEE International Conference on Robotics and Automation (ICRA) (IEEE, 2014)*, pp. 6808–6815.
89. R. Mir, P. Ojaghi, A. Marjaninejad, M. Wehner, F. Valero-Cuevas, “Active sensing in a bioinspired hand as an enabler of implicit curriculum learning for manipulation” in *The 9.5th International Symposium on Adaptive Motion of Animals and Machines (AMAM2021) (Springer, 2021)*, p. 61.
90. V. Kumar, E. Todorov, S. Levine, “Optimal control with learned local models: Application to dexterous manipulation” in *2016 IEEE International Conference on Robotics and Automation (ICRA) (IEEE, 2016)*, pp. 378–383.
91. J. Wang, Y. Liu, B. Li, Reinforcement learning with perturbed rewards. *Proc. AAAI Conf. Artif. Intell.* **34**, 6202–6209 (2020).
92. C.-Y. Tang, C.-H. Liu, W.-K. Chen, S. D. You, Implementing action mask in proximal policy optimization (ppo) algorithm. *ICT Express* **6**, 200–203 (2020).
93. P. Dhariwal, C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford, J. Schulman, S. Sidor, Y. Wu, P. Zhokhov, Openai baselines, 2017; <https://github.com/openai/baselines>.
94. K. Mills, P. Ronagh, I. Tambllyn, Finding the ground state of spin Hamiltonians with reinforcement learning. *Nat. Mach. Intell.* **2**, 509–517 (2020).
95. A. Gotmare, N. S. Keskar, C. Xiong, R. Socher, A closer look at deep learning heuristics: Learning rate restarts, warmup and distillation. arXiv:1810.13243 [cs.LG] (2018).
96. Y. Xiong, L.-C. Lan, X. Chen, R. Wang, C.-J. Hsieh, “Learning to schedule learning rate with graph neural networks” in *International Conference on Learning Representation (ICLR) (OpenReview, 2022)*, pp. 1–21.
97. A. Senior, G. Heigold, M. Ranzato, K. Yang, “An empirical study of learning rates in deep neural networks for speech recognition” in *2013 IEEE international conference on acoustics, speech and signal processing (IEEE, 2013)*, pp. 6724–6728.
98. I. Loshchilov, F. Hutter, Sgdr: Stochastic gradient descent with warm restarts. arXiv:1608.03983 [cs.LG] (2016).

#### Acknowledgments

**Funding:** This work is supported in part by the NIH R21 NS113613-01A1, NSF 2113096 CRCNS US-Japan, DOD CDMRP Grant MR150091, and DARPA-L2M W911NF1820264 awarded to F.J.V.-C. and the USC Viterbi School of Engineering Fellowships to A.M. and R.M. This work does not necessarily represent the views of the NIH, NSF, DoD, or DARPA. This work was supported in part by the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education funded by the Wisconsin Alumni Research Foundation awarded to M.W. **Author contributions:** All authors contributed to writing—review and editing. P.O., R.M., A.M., A.E., and F.J.V.-C. contributed with writing—original draft, formal analysis, and visualization. P.O., R.M., A.M., M.W., and F.J.V.-C. contributed to conceptualization. P.O., R.M., and A.M. contributed to software and investigation and worked on data curation and its validation. P.O., R.M., A.M., and F.J.V.-C. contributed to project administration, methodology, and resources. A.M., M.W., and F.J.V.-C. did the supervision of the team. A.M. and F.J.V.-C. contributed to funding acquisition. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data and code needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials and are available in the following repository: [github.com/pojaghi/In-hand-manipulation](https://github.com/pojaghi/In-hand-manipulation) (permanent repository: <https://doi.org/10.7910/DVN/P6200E>).

Submitted 16 April 2024

Accepted 26 February 2025

Published 2 April 2025

10.1126/sciadv.adp8407