Completion of the Infeasible Actions of Others: Goal Inference by Dynamical Invariant

Takuma Torii
tak.torii@jaist.ac.jp
Shohei Hidaka
shhidaka@jaist.ac.jp
Japan Advanced Institute of Science and Technology,
Nomi, Ishikawa 923-1211, Japan

To help another person, we need to infer his or her goal and intention and then perform the action that he or she was unable to perform to meet the intended goal. In this study, we investigate a computational mechanism for inferring someone's intention and goal from that person's incomplete action to enable the action to be completed on his or her behalf. As a minimal and idealized motor control task of this type, we analyzed single-link pendulum control tasks by manipulating the underlying goals. By analyzing behaviors generated by multiple types of these tasks, we found that a type of fractal dimension of movements is characteristic of the difference in the underlying motor controllers, which reflect the difference in the underlying goals. To test whether an incomplete action can be completed using this property of the action trajectory, we demonstrated that the simulated pendulum controller can perform an action in the direction of the underlying goal by using the fractal dimension as a criterion for similarity in movements.

1 Introduction: Imitation of Action

As a method of social learning from others, children imitate their parents' movements in early development (Meltzoff, 1995). Imitation, as a behavioral basis for understanding other's goals and intentions, is considered a mechanism for preserving social and cultural knowledge. From the perspective of cultural evolution, it plays a key role as a "latchet," which preserves the skills and knowledge obtained by our ancestors, preventing human cultural knowledge from moving backward (Tomasello, 2001). Teaching techniques that show and mimic a demonstration are commonly adopted in not only education but also robot learning (Schaal, 1999).

In a typical imitation (Breazeal & Scassellati, 2002), a demonstrator (e.g., parent) shows the imitator (e.g., child) an action with an intention. In this letter, we employ the definition of intention and action described by Bernstein (1996): *intention* means either motor planning or motor control to

Neural Computation **33, 2996–3026** (2021)
https://doi.org/10.1162/neco_a_01437

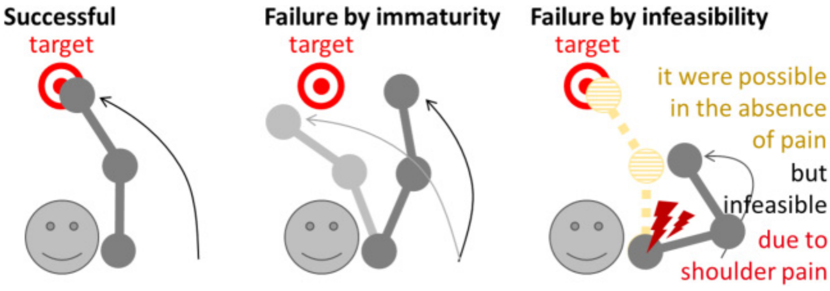

Figure 1: Two types of failure illustrated using the reaching task. Successful: The actor accurately reaches the target. Failure by immaturity: The actor is unable to accurately reach the target due to poor control. Failure by infeasibility: The actor is unable to reach the target due to the presence of pain. The actor would be able to reach the target if he or she had no pain.

achieve a certain goal, leading to a series of choices or a movement toward a certain goal, and *action* means a movement with an intention to achieve a certain goal. Using these terms, we operationally define the "success" or optimality of an action as consistency between the generated movement and the goal of the motor control system, and define the "failure" of an action as inconsistency between them. In this definition, we differentiate the decision to use motor control from actual task performance and define the success or failure of an action in terms of the former. We further refine our definition of the failure of actions for two distinct causes below (illustrated in Figure 1).

Consider two scenarios in which a human demonstration can end in failure without completing a given task. The actor intended to reach a target but could not do so because of inaccurate motor control or could not move her arm freely due to an injury.

Although the action ends in failure in both scenarios, the reasons underlying why the actor failed to complete the given task are totally different. In scenario 1 (see Figure 1, failure by immaturity), the reason for the failure is an insufficiency in the accuracy of the actor's motor control for the given task, which can likely be overcome with additional motor learning. However, in scenario 2 (see Figure 1, failure by infeasibility), the reason for the failure is temporary or permanent inability to perform an appropriate action. By our definition, the action in scenario 2 is considered a hypothetical success because the actor's motor control is optimal for the task in question and her action would be successful if she were not injured. In this study, we refer to the former type of failure as "failure by immaturity" and the latter as "failure by infeasibility."

While problems closely related to imitation have been studied in robotics, the majority of these have been classified as failure by immaturity.

For example, Schaal proposed a scheme called learning from demonstration (LFD; Schaal, 1997, 1999), which aimed to utilize human demonstration to initialize and improve a robot controller. Typically, combined with reinforcement learning (Sutton & Barto, 1998; Doya, 1999), a set of human-demonstrated trajectories is used to provide an initial guess of controller parameters such as the Q function to leverage learning (Schaal, 1997) and/or update trained controllers based on human performance evaluation (Argall, Browning, & Veloso, 2007). The idea of LFD has been extended to inverse reinforcement learning (IRL; Ng & Russell, 2000; Abbeel & Ng, 2004), whose aim is to infer the unknown reward function given a task structure (states, actions, and environment) and a set of trajectories produced by experts of the given task. The recent development of IRL algorithms (Ziebart, Maas, Bagnell, & Dey, 2008; Babes-Vroman, Marivate, Subramanian, & Littman, 2011; Ho, Littman, MacGlashan, Cushman, & Austerweil, 2016) has led to great success in such applications.

LFD and IRL studies have typically examined certain tasks under the failure-by-immaturity class. Both LFD and IRL typically assume that all given action trajectories are successful or fall under the failure-by-immaturity class with respect to an unknown but fixed task. Recent studies (Grollman & Billard, 2011; Shiarlis, Messias, & Whiteson, 2016) on learning from "failed" demonstrations (reviewed in Zhifei & Joo, 2012) have attempted to utilize nonexpert trajectories in failure-by-immaturity.

In detail, Grollman and Billard (2011) studied how a robot can learn only from nonexpert demonstrations of failure by immaturity. Namely, they assumed that the provided "failed" demonstrations are distributed around the "successful" trajectory in space. In this sense, they assumed that the failed trajectories were similar to the successful ones and on average contained information about the successful ones. A subsequent study found that a handcrafted reward function for the target task or human performance evaluation is required to train an acceptable robot controller (Grollman & Billard, 2012). Recently, Shiarlis et al. (2016) studied how nonexpert demonstrations can be used to improve the performance of existing IRL algorithms. Shiarlis et al. (2016) utilized nonexpert demonstrations of failure by immaturity as auxiliary information and successful demonstrations to train IRL systems as a set of positive and negative samples, respectively, to behave more like successful ones and less like failed ones. Shiarlis et al. (2016) showed that successful demonstrations must be provided to facilitate learning of a controller from demonstrations.

Most previous studies on learning from (failed) demonstrations assumed that the demonstration was either successful or a case of failure by immaturity rather than failure by infeasiblity. Thus, the learner or imitator in this scheme can access an optimal or near-optimal (with noise) demonstration. However, children, even in early development, can go further: they can learn and complete a failed demonstration by infeasibility, which has no action at its goal state. In this case, what the imitator has to infer is the goal

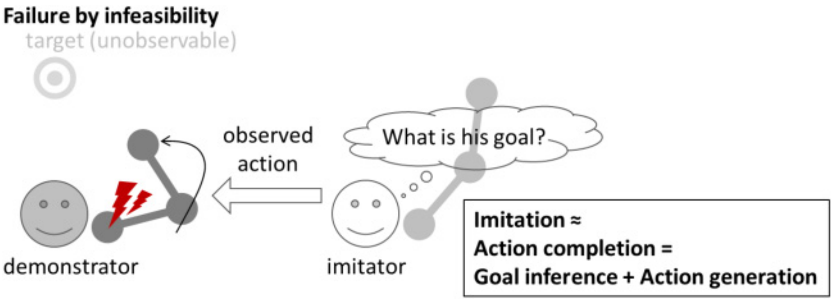

Figure 2: Problem setting for this study: goal inference and action generation from an action that failed by infeasibility. This is a model of a typical situation in which the demonstrator needs the help of the imitator. To help the demonstrator, the imitator has to infer the goal of the demonstrator and complete the demonstrator's action that failed by infeasibility.

underlying the demonstrator's nonoptimal movement, in which the goal state is absent. We think that failure by infeasibility (scenario 2) is needed to help others—that is, it is necessary to complete an action in the course of meeting its goal without knowing the goal state. Relevant developmental studies have shown that two-year-olds can "help" others by completing their action (Meltzoff, 1995; Warneken & Tomasello, 2006).

In this study, we examine a computational mechanism for imitation learning from an action that failed by infeasibility, where the imitator does not know the demonstrator's goal or the intention behind his action (as illustrated in Figure 2). Given our proposed setting, imitation learning requires solving the following two major classes of problems: (1) *identification of action features*, in which two actions with different intentions can be discriminated, and (2) *completion of observed action*, in which an incomplete part of someone's action to meet a goal is extrapolated and an appropriate action to meet that goal is performed. The first problem, identification of features, requires identifying features correlated with the intentional difference (functional difference in motor control) behind the observed action rather than features that just describe apparent movements. Inferring an intention or goal behind an action is, however, generally an ill-posed problem: a pair of similar movements can be produced by two very different intentions and/or with two different goals. The second problem, action completion, requires not only the identification of features but also ensuring that the imitator's own action meets an inferred goal for some observed portion of the demonstrator's incomplete action that the demonstrator intended but failed to complete.

In this letter, we aim to address the two problems we posed above in imitation learning: (1) action recognition and (2) action completion by performing a numerical study on a task that involves controlling a physical object—a single-stick pendulum. We suppose that this simple control task is minimally sufficient to capture the essential aspects of goal imitation: how one recognizes the intention (motor control) behind a given action and how one performs the action. The primary objective of this letter is to provide computational proofs-of-concept for our hypothesis that some degrees-of-freedom (DoF) is critical for characterizing the underlying goal and intention of an observed action (described in the next section). Therefore, in the computer simulation studies in this letter, we suppose that the imitator can obtain sufficient trajectory data from the demonstrator to learn action features. This allows us to explore the primary problem, the principle of the computational possibility of goal inference, separately from other technical problems, such as learning from a small training data set. This assumption may be a limitation of our study and is discussed in section 5.

Although the control task involving a pendulum may be considered overly simple in its structural complexity compared to the human body, we think this task has very similar characteristics to the experimental task reported by Warneken and Tomasello (2006). In their experiment, Warneken and Tomasello exposed 18-month-old children to an adult (experimenter)'s goal-failed action and investigated whether these children could infer the adult's latent goal, which was not demonstrated, and could help the adult complete the goal-failed action. The research suggested that children of this age can infer others' goals and complete others' actions.

In principle, children in such an experiment are required to (1) recognize the adult's failed goal and/or intention and (2) perform their own action by controlling their own body to meet the adult's goal. In this letter, tasks 1 and 2 are, respectively, called *recognition* and *completion* tasks for goal imitation. We illustrate how our simulation framework captures the goal imitation behavior and report two simulation studies for our recognition and completion task.

2 Simulation Design

2.1 A Situation That Requires Recognition and Completion of Other's Action. First, we briefly introduce the psychological experiment performed by Warneken and Tomasello (2006) (abbreviated as WT hereafter) as a representative situation against which we modeled our theoretical framework. WT investigated whether children can infer a demonstrator's goal and the intention behind their behavior. In the experimental (goal-failed) condition, called out-of-reach, the demonstrator accidentally dropped a marker on the floor and was unable to reach for it. In the control (goal-achieved) condition, the demonstrator intentionally dropped a marker on the floor. The former condition implicitly calls for the child to help the demonstrator achieve her unsuccessful intention/goal, namely, to pick up the marker, while the latter does not. The experimental and control conditions were designed such that the demonstrator's apparent bodily movements were similar (e.g., both dropped a marker), whereas the underlying intention/goal behind the action was different. WT showed that the children more frequently showed helping behaviors in the experimental condition than the control condition.

2.2 A Model of Recognition and Completion of Other's Action. In this study, we designed a simulation framework to capture the essence of WT's experimental design in minimal form. Specifically, we employed a single-link pendulum as a simplified human body. The imitator (i.e., the hypothetical child) and the demonstrator must both control a pendulum to perform an action (i.e., a goal-directed movement). The demonstrator's goal is to keep the body of the pendulum at the top-most position of its trajectory (opposite to gravity) as much as possible subject to given bodily constraints, a set of physical parameters for the pendulum (e.g., mass and length). The demonstrator's intention is motor control (or policy in terms of reinforcement learning) of the pendulum, which gives angular acceleration (force) as a function of the angle and angular velocity of the pendulum. An action of the demonstrator is to manipulate the movement (trajectory) of the pendulum, as represented by either an orbit of the (x, y) coordinates or a vector of the angle and angular velocity, which are generated using a given pair of initial conditions and the demonstrator's controller.

We hypothesize that the essential difference between the experimental (goal-failed) and control (goal-achieved) conditions in the study conducted by WT is captured by the degree of optimality of the intention and action with respect to the given goal. Suppose there are controllers A and B, which are optimal for the distinct goals GA and GB, respectively. If the demonstrator uses controller A for goal GA, the generated movement would be optimal and considered a successful action. In contrast, if the demonstrator uses controller B for goal GA, the generated movement would in general be suboptimal and would be considered a failed action. We consider the former case to be analogous to the control (goal-achieved) condition in the experiment conducted by WT in which a successful action was performed and did not lead the child to help the demonstrator and the latter to be analogous to the experimental (goal-failed) condition that led the child to help the demonstrator.

Accordingly, we designed two tasks for demonstrators with different combinations of task goals and constraints. We called the first task (A) the *swing-up task*, in which the goal is to keep the mass of the pendulum as close as possible to the top of the angle space without any obstruction (see Figure 3A). The goal is implicitly defined by the maximum of the reward function $r(\theta) = \cos\theta$ of the angle $\theta \in \mathbb{R}$, which takes the maximum at the top-most position $\theta = 2\ n\pi$ for any $n \in \mathbb{Z}$ in this angle coordination. We called the second task (B) the *swing-up-no-hit* task, in which the movement of the pendulum is constrained within a given angle range, called the *feasible angle space*. The remaining angle space, in which the pendulum is not

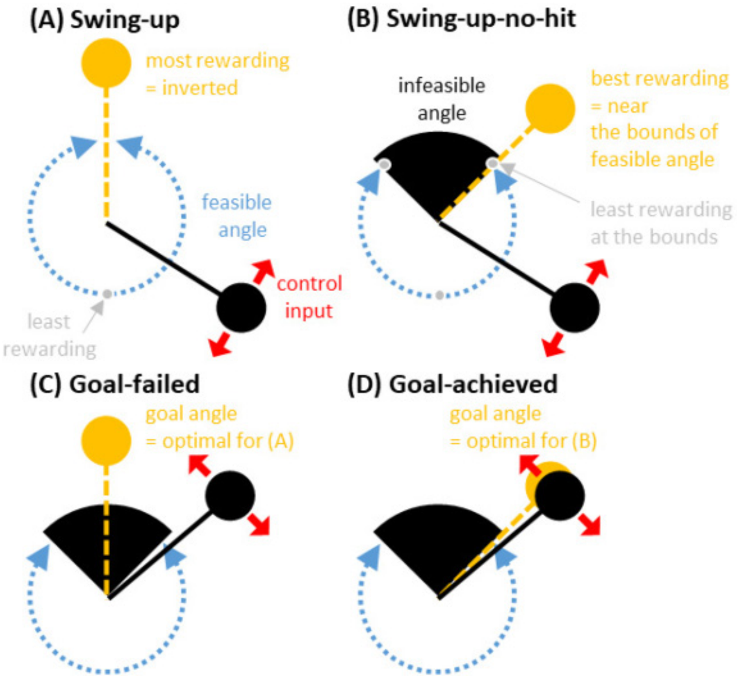

Figure 3: Simulation design analogous to experimental tasks in Warneken and Tomasello (2006). (A) The swing-up task. The most rewarding angle is when the mass of the pendulum is at the top-most position of the angle space ($\theta = 0$), and the least rewarding angle is at the bottom of the angle space ($\theta = \pi$). (B) The swing-up-no-hit task. The least rewarding angles are when the mass of the pendulum is at the bottom of the angle space ($\theta = \pi$) and at the bounds of the infeasible region (black: $\theta = \pm\pi/8$). The most rewarding angle is somewhere close to the top of the angle space within the feasible region. It is optimal to keep the pendulum swinging without touching the bounds. (C) Goal-failed demonstration: performing task B with the control that is optimal for task A. (D) Goal-achieved demonstration: performing task B with the control that is optimal for task B.

allowed to enter in the swing-up-no-hit task, is called the infeasible angle space. The black region in Figure 3B shows the infeasible angle space. The goal of the swing-up-no-hit task is to keep the mass of the pendulum as close as possible to the top of the angle space while remaining within the feasible angle space. In the swing-up-no-hit task (B), the demonstrator will be given the least reward $r(\theta) = -1 = r(\pi)$ for any θ in the infeasible angle space including its boundary; otherwise, the demonstrator is given at each time step the reward $r(\theta) = \cos\theta$ as a function of angle with the least value $r(\pi) = -1$. The degree of optimality (i.e., numerical indicator of match or mismatch between the goal of the task and the action) is defined as the cumulative sum of rewards for a given action trajectory over time relative to the largest cumulative sum of the theoretically best action trajectory for the given task.

Given these two types of tasks, we defined the *goal-failed* condition as a mismatch between the task and action in which the demonstrator is performing the swing-up-no-hit task (with the infeasible region) by controlling the pendulum using a controller that is optimal for the swing-up task (see Figure 3C). We consider this goal-failed condition to be analogous to the experimental condition in the study conducted by WT in which a failed action was demonstrated. We also defined the *goal-achieved* condition, which we consider to be analogous to WT's control condition, as a match between the task and action in which the demonstrator is performing the swing-up-no-hit task by controlling the pendulum using a controller that is optimal for the swing-up-no-hit task (see Figure 3D). We expect that the demonstrator in our goal-failed condition (see Figure 3C), but not the goal-achieved condition (see Figure 3D), will perform an action that is suboptimal for the swing-up-no-hit task, which may appear similar on the surface but is essentially different from the action intended to be optimal for the swing-up task.

The imitator, in turn, observes two types of goal-failed and goal-achieved actions, which are potentially different, and analyzes their potential difference based only on the observed action trajectories. This situation corresponds to simulation I (see section 3), in which we investigated recognition of the potential difference between goal-failed and goal-achieved actions.

After some visual inspection of actions, the imitator is expected to perform his or her own actions to complete the demonstrator's action (i.e., "help" the demonstrator) if the action is incomplete or goal-failed. This situation corresponds to simulation II (see section 4) in which we investigated action generation based on observation of the demonstrator's incomplete or goal-failed actions.

2.3 Pendulum Control. A mathematical model of a simple pendulum is composed of a link of length $l = 1$ with one end fixed at the origin and a point mass $m = 1$ at the other end. The state of this pendulum is identified by the angle $\theta \in \mathbb{R}$ of the link, relative to angle zero, which corresponds to the top-most position of the angle space, and the angular velocity $\dot{\theta} \in \mathbb{R}$, the first-order time derivative of the angle. The equation of motion is given by

$$ml^2\ddot{\theta} - mgl\sin\theta = f(\theta, \dot{\theta}) + \varepsilon, \tag{2.1}$$

where $g = 9.8$ is the gravity constant, $f(\theta, \dot{\theta})$ is the state-dependent control input (torque) from a controller f, and $\varepsilon \sim N(0, \sigma)$ is the intrinsic noise of the system, time independently sampled from a normal distribution with variance σ^2.

The pendulum swing-up task is classically used in feedback control theory (Doya, 1999), originally used to design a controller f that can swing the pendulum and maintain it at about the top-most position where $\theta = 0$. The controller for this task is defined by the function f, which outputs torque $f(\theta, \dot{\theta}) \in \mathbb{R}$ for any given state $(\theta, \dot{\theta}) \in (-\pi, \pi] \times (-2\pi, 2\pi]$. The goal of the task is implicitly and quantitatively represented by the reward function r (see equation 2.4 or 2.5). With this reward function, we can define the goal-meeting action as an action with the maximal reward value (or large enough to be considered an approximation of the maximum) $\sum_t r(\theta_t)$ as a function of the controller (see the next section for details). In each run of the simulation, the initial position of the pendulum was set such that $\dot{\theta} = 0$ and angle θ drawn from the uniform distribution ranged by $\theta \in \pm[\pi/8, \pi)$.

2.4 Energy-Based Swing-Up Controller. The simple pendulum defined in the previous section is well characterized by the mechanical energy of the system—that is, the sum of the kinetic and potential energy:

$$E(\theta, \dot{\theta}) = \frac{1}{2} m l^2 \dot{\theta}^2 + mgl\,(\cos\theta - 1). \tag{2.2}$$

In the pendulum swing-up task, its goal state or the most rewarding state (i.e., the pendulum is inverted with the mass at the top-most position $\theta = 0$ with zero velocity $\dot{\theta} = 0$) corresponds to the state with energy $E(0, 0) = 0$. The simple pendulum preserves the mechanical energy over time in this state if the control input is set to $u = 0$ without any noise ($\sigma = 0$). Thus, one way to meet the goal of the swing-up task is to keep the mechanical energy at zero ($E(\theta, \dot{\theta}) = 0$). Based on this observation, Astrom and Furuta (2000) defined the energy-based controller of the simple pendulum as

$$f_G(\theta, \dot{\theta}) = -(E(\theta, \dot{\theta}) - E(G, 0))\,\dot{\theta}, \tag{2.3}$$

with which one can reduce the difference between the energy $E(\theta, \dot{\theta})$ of the current state and the target energy $E(G, 0)$ with the goal angle $G \in \mathbb{R}$. In our simulation, we employed this energy-based controller to generate an action toward the goal angle G.

In our previous work (Torii & Hidaka, 2017), we adopted a controller (or policy) based on reinforcement learning to study the action recognition task described in section 3. We obtained the same qualitative results to those in this letter. To study the action completion task described in section 4, we adopted the energy-based controller throughout the study, whose scalar parametric form is very convenient compared to reinforcement learning, which requires computationally expensive training of the policy function.

2.5 Goal-Achieved and Goal-Failed Action. For each of the demonstrators in the swing-up and swing-up-no-hit tasks, two different reward functions $r(\theta)$ are assumed as described in section 2.2. In the swing-up task, the reward function is

$$r(\theta) = \cos\theta, \tag{2.4}$$

which indicates that the top-most position $\theta = 0$ is the most rewarding position for the pendulum. In the swing-up-no-hit task with an infeasible angle space of $[-\theta_{\min}, +\theta_{\min}]$, the reward function is

$$r_{\theta_{\min}}(\theta) = \begin{cases} \cos(\pi) = -1 & \text{if } \theta \in [-\theta_{\min}, +\theta_{\min}] \\ \cos\theta & \text{otherwise} \end{cases}. \tag{2.5}$$

Specifically, we set $\theta_{\min} = \pi/8$ in the swing-up-no-hit task. The optimal controllers are different for the two tasks with the different reward functions r and $r_{\theta_{\min}}$. The optimal energy-based controller for the swing-up task and the swing-up-no-hit task is the controller (see equation 2.3) with the goal angle $G = 0$ and $G = \theta_{\min}$, respectively.

For both the swing-up and swing-up-no-hit tasks, we applied the energy-based controller (Astrom & Furuta, 2000) with some goal angle G introduced in the previous section. Since the energy-based controller with the goal angle $G = 0$ was originally designed for the pendulum swing-up task with no angle constraint, this energy-based controller is not optimal for the swing-up-no-hit task with the constrained pendulum: it does not supply sufficient torque to hold the pendulum against gravity. As a result, it produces a repeated swinging movement, unlike the behavior without the infeasible boundary, in which it holds the pendulum still at the goal angle.

For visual inspection of the movements generated by these two distinct controllers with $G = 0$ or $G = \theta_{\min}$, Figure 4 shows the two typical angular time series generated by these controllers. Figure 4 (top) shows the typical actions (time series of angles) performed by the goal-failed demonstrator with the goal angle $G = 0$ in the swing-up task (see Figure 3C). Figure 4 (bottom) shows the typical actions performed by the goal-achieved demonstrator with the goal angle $G = \theta_{\min}$ in the swing-up-no-hit task (see Figure 3D). Both movements result in swinging within the range $\theta \in \pm[\max\{G, \theta_{\min}\}, \pi]$, despite having different goal angles and feasible angle spaces. These movements look similar in their angle dynamics, which simulate the movement similarity in the experiment by WT (e.g., dropping a marker accidentally and intentionally). However, as shown in Figure 4, their mechanical energy, a direct indicator of their controller, can reveal differences between the two actions.

When the pendulum collides with the bounds, some loss of mechanical energy occurs because the height of the pendulum forcibly remains

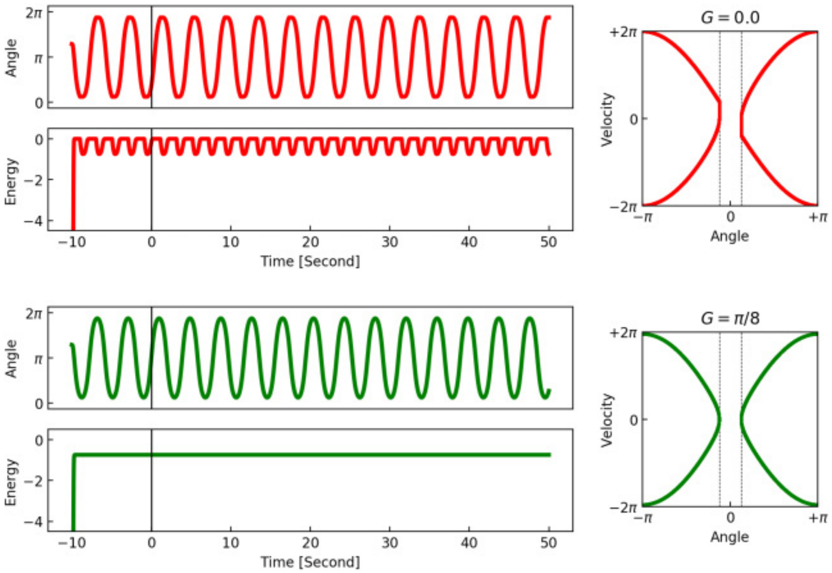

Figure 4: A typical time series of angle and mechanical energy generated by the goal-failed demonstrator (top panels) with $G = 0$ and the goal-achieved demonstrator (bottom panels) with $G = \theta_{\mathrm{min}}$. The dashed lines in the right panels indicate the bound $\theta_{\mathrm{min}} = \pi/8$.

unchanged and the body decelerates to an angular velocity of zero, which can be visually observed in both the energy-time series and the trajectory in the angle-velocity plane in the top panel of Figure 4. In contrast, no such loss of energy is observed in the bottom panel because the pendulum rarely collides with the bounds.

2.6 Features for Detecting Intentional Differences. According to our definition of the goal-failed and goal-achieved conditions, the intention behind a movement that is optimal for the swing-up task does not match that for the swing-up-no-hit task (see Figure 3C). Other than in this particular case, many other actions that fail by infeasibility, including those in the study by WT, essentially display this type of mismatch between some originally intended task and the actual performed task. One of the critical features common to these types of tasks is that the task to work has an additional unexpected obstacle that is absent in the original task, for which the controller is optimal.

Beyond specific differences across different tasks, we hypothesize that these types of failures may be characterized by the existence of some additional factor complicating the originally intended task. In WT's condition in which a marker was accidentally dropped, the demonstrator was not ready for the situation in which he or she was required to pick up the dropped marker; the accidental dropping of the marker introduces an additional complexity to the originally intended task: to carry the marker to some location (without dropping it). This is analogous to the goal-failed condition in our pendulum simulation: the additional obstacle, the limitation in the feasible angle, causes suboptimality of the original motor control in this unexpected new task.

What characteristics can be used to detect such suboptimality in an action? In this study, we hypothesize that this additional factor of complexity can be detected in the degrees-of-freedom (DoF) of the given system.

Let us consider a successful action, for example, the goal-achieved condition of the pendulum control task. Such an action is expected to flow smoothly, without any sudden change in its motion trajectory. Thus, the movement can be closely approximated using a set of differential equations with a relatively small number of variables. In contrast, an action that fails by infeasibility, for example, the goal-failed condition of the pendulum control task, is expected to have some discontinuous or nonsmooth change in its motion trajectory, such as at the time point before or after an unexpected accident for the given system. Thus, before and after this change, such a system would be better described using two or more distinct sets of differential equations.

Although it is technically difficult to identify such differences in the underlying systems (or sets of differential equations) in full detail here, it should be clear that the underlying controller in these two cases would differ in their DoF. This consideration leads us to the hypothesis that some difference in the DoF of movement is diagnostic of successful and failed actions.

In this letter, we specifically employ a type of fractal dimension, called pointwise dimension (see section 3.1), of the actions as an indicator of the DoF of the underlying controller and test whether it is characteristic of the difference in intention underlying the actions. In the following two sections, we examine our hypothesis by analyzing the movement data generated by the simulated pendulum control task. We divided our analyses into action recognition and action completion.

First, we analyzed the recognition task from the imitator's perspective by examining which features of the movements the imitator (observer of the actions) was able to discriminate between the goal-achieved and goal-failed actions. Success recognition, the ability to tell the difference between two qualitatively different actions, is considered necessary to complete another's failed action.

Second, we analyzed the completion task by asking whether the characteristic features of the intention underlying actions, as identified in the first analysis, are sufficient to generate an action to complete a goal-failed action. Here, completion of the action means that the imitator performs an action that meets the goal that an observed demonstrator’s action failed to meet. As such, a goal-failed action is incomplete by definition and not fully observed by the imitator; thus, the imitator is required to extrapolate the observed action to generate the originally intended action. This action completion task needs not just recognition of the qualitative difference in actions but also some identification of the demonstrator’s failed action and the imitator’s action.

3 Simulation I: Action Recognition Task

In simulation I, we investigate whether the imitator can tell the difference between the two different intentions underlying the actions performed by the demonstrators in the goal-failed and goal-achieved conditions. The goal of this simulation is to analyze and identify the feature that is most characteristic of the latent intention of actions.

Specifically, we listed several features typically used in time series analysis, such as angle (or angular position), angular velocity, angular acceleration, angular jerk, power spectrum, mechanical energy, and pointwise dimension. We hypothesized that pointwise dimension would be most characteristic of the latent intention of actions for this analysis. Angle, angular velocity, and power spectrum are commonly employed features of movements in the literature. They are also fitting for our simulation, as motor control is a function of angle and angular velocity, and the generated movement is periodic. Mechanical energy is the very concept defining the motor control task (see equation 2.3), and we thus expect mechanical energy to be the best possible feature in theory to characterize the intention (motor control). However, a naive imitator, such as a child, who is ignorant of the demonstrator’s physical properties may not have direct access to the mechanical energy because of the need for knowledge of the physical parameters of the pendulum (i.e., mass m and length l in equation 2.1, which are necessary to compute the mechanical energy of the pendulum system). Thus, we treated mechanical energy as an indicator of the best possible (but unlikely to be directly accessible) reference feature for the recognition task in our analysis.

Finally, given that the pointwise dimension indicates the latent DoF of an underlying dynamical system, we hypothesize that it is an indicator of task-system complexity and is characteristic of the intentional difference between movements performed in the goal-failed and goal-achieved conditions. We tested this hypothesis by evaluating recognition performance using the pointwise dimension compared to that using the reference feature: mechanical energy in the classification of movements with different intentions.

3.1 Pointwise Dimensions.

To characterize complexity in each demonstrator’s movements, we analyzed the attractor dimension of the movements by treating it as a dynamical system. Specifically, we exploited a type of fractal dimension called pointwise dimension for classification analysis. The pointwise dimension is a type of dimension defined for a small, open set or measure in the set, including a point in a given set (see Cutler, 1993; Young, 1982, for details). Formally, for a set of points X in a topological space $\mathbb{R}^n$, the pointwise dimension $d(x)$ at point $x \in X$ is defined (if it exists) by the local scaling exponent of the associated probability measure μ on X such that $\mu(B(x, \epsilon)) \sim \epsilon^{d(x)}$ as $\epsilon \to 0$, commonly expressed as

$$d(x) = \lim_{\epsilon \to 0} \frac{\log \mu(B(x, \epsilon))}{\log \epsilon}, \tag{3.1}$$

where $B(x, \epsilon) \subseteq X$ gives a subset of points around x within distance ϵ. Pointwise dimension is invariant under arbitrary smooth transformation. As it is associated with each point, we can analyze the distribution of the pointwise dimension across points. Informally speaking, the pointwise dimension of a point characterizes the measurable space of a certain dimension that surrounds the point. We have developed a statistical technique to estimate the pointwise dimension for a set of data points (Hidaka & Kashyap, 2013). Using this technique, each point in the data set is assigned a positive value of pointwise dimension.

3.2 Two-Class Classification. We performed classification analyses of demonstrator types based on each of the features described. Performance of the classification is used as a measure of how well each feature discriminates among demonstrator types. Specifically, for this two-class classification task, the imitator is exposed to a time series of a pair of angles that reflect each movement demonstrated in the goal-achieved and goal-failed conditions. Part of each time series corresponding to the first 10 seconds of the task was excluded from the training data because these were transient periods that were heavily dependent on the initial state. The rest of the time series, corresponding to the last 50 seconds of the movement (of 5000 sample points), was used as the training data for classification. We used a single, long time series because the system is expected to be ergodic, defined as a time series with any initial starting state that eventually converges to the same stationary near-periodic dynamical system (with some intrinsic noise in motor control).

For classification, we considered the following features: angle (or angular position), angular velocity, angular acceleration, angular jerk (the third derivative of angle), power spectrum, mechanical energy, and pointwise dimension. Given a time series of angle (or angular position), the time series of angular velocity, acceleration, and jerk was calculated by taking the first-, second-, and third-order difference of the angle time series. The third derivative of the position, called "jerk," is a notable feature that is hypothesized to be critical in the minimum jerk and/or minimum torque-change trajectory for human motor control of reaching (Hogan, 1984; Uno, Kawato, & Suzuki, 1989). The data points for the power spectrum feature were constructed as a collection of frequencies with the largest powers in the power spectrum of angles computed within a moving time window size of 5 seconds. Details of the construction of the pointwise dimension feature are described later. In contrast to the features described, which can only be computed using the observable time series, computation of mechanical energy requires knowledge of the physical properties, such as the body mass and length, of the pendulum system, as evident from equation 2.2.

To analyze the degree of contribution of pointwise dimension to recognizing the underlying controller of the system, the pointwise dimension associated with each data point was estimated from a time series of coordinate values $(x_t, y_t) = (\sin\theta_t, \cos\theta_t)$ of the pendulum. As pointwise dimension is an invariant under arbitrary smooth transformation, we obtained essentially the same estimate as that from the time series of angle. In dynamical systems theory, Takens's embedding theorem states that a diffeomorphism of a smooth attractor in a latent high-dimensional space can be reconstructed from a univariate or low-dimensional time series, which is a projection on a subset of the original high-dimensional state space, using time-delay coordinates with a sufficiently high dimension for embedding the attractor. Therefore, the positional time series $\{(x_t, y_t)\}_t$ was first embedded into the time-delay coordinates of the embedding dimension $2k$, $\{(x_t, y_t, x_{t+1}, y_{t+1}, \ldots, x_{t+k-1}, y_{t+k-1})\}_t$. Then the embedded $2k$-dimensional time series (of $5000 - k + 1$ sample points) was used to estimate the pointwise dimension, equation 3.1, for the time series. We mostly adopted $k = 20$ for the pendulum system with a controller.

For the recognition (two-class classification) task, the feature points in a feature time series are treated as independent samples. The classification procedure is illustrated in Figure 5. X_F and X'_F denote two distinct sets of feature points in the training and test data set, respectively, which are constructed from distinct actions generated by the goal-failed demonstrator. Similarly, X_A and X'_A denote those of the goal-achieved demonstrator. For features other than those for pointwise dimension, given a set of feature points as training data, either X_F or X_A, we used the gaussian mixture model, in which each class of data is distributed as one or more multivariate normal distribution(s) over a given feature space. We chose the gaussian mixture model because of its computational simplicity for constructing two sample probability functions of a variable (i.e., a feature). We denoted the probability density function of a certain feature x estimated with the goal-failed action(s) X_F as $p_F(x)$ and that estimated with goal-achieved action(s) X_A as $p_A(x)$. Using these sample probability density functions, the imitator asserts that a given test feature $x \in X'_F \cup X'_A$ belongs to the goal-failed demonstrator if $p_F(x) > p_A(x)$; otherwise, it belongs to the goal-achieved demonstrator. The classification accuracy is defined as the proportion of

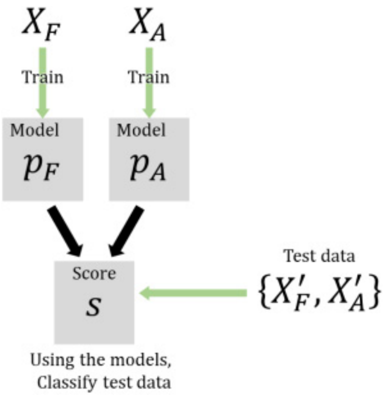

Figure 5: Procedure for the recognition task.

correct responses, which is defined by the equality between the underlying class and the asserted class for each unit of a given time series. Precisely, the correct response was defined by $p_F(x) > p_A(x)$ if a test feature x was indeed sampled from $x \in X'_F$; otherwise, namely, $x \in X'_A$, the correct response was $p_F(x) < p_A(x)$. For each feature, we reported the classification accuracy of the gaussian mixture model, with the number of multivariate normal distributions in the gaussian mixture model selected based on the minimum Akaike information criterion (Akaike, 1974).

To perform the recognition task using pointwise dimension as a feature, we used the statistical model underlying the dimension estimation method proposed by Hidaka and Kashyap (2013). The method or dimension estimator constructs a model for given data as a mixture of multiple Weibull-gamma distributions, each of which has an associated parameter representing the fractal dimension. This method can also be used to calculate the probability that a sample data point x with time-delay embedding belongs to the mixture model. Therefore, for the probability density functions of the pointwise dimension feature, we adopted the Weibull-gamma mixture model rather than the gaussian mixture model used for the other features. The training and test data for, say, X_F and X'_F, were both obtained by time-delay embedding the positional time series. For this recognition task, the spatial neighborhood of a sample point $x \in X'_F \cup X'_A$ was calculated within the feature space spanned by training data X_F (or X_A) associated with the probability density function p_F (or p_A). The number of mixture components of the Weibull-gamma mixture model was selected based on the minimum Akaike information criterion (Akaike, 1974).

3.3 Classification Results. Figure 6 shows the classification accuracy for each feature of the test data set. As both the training and test data contain

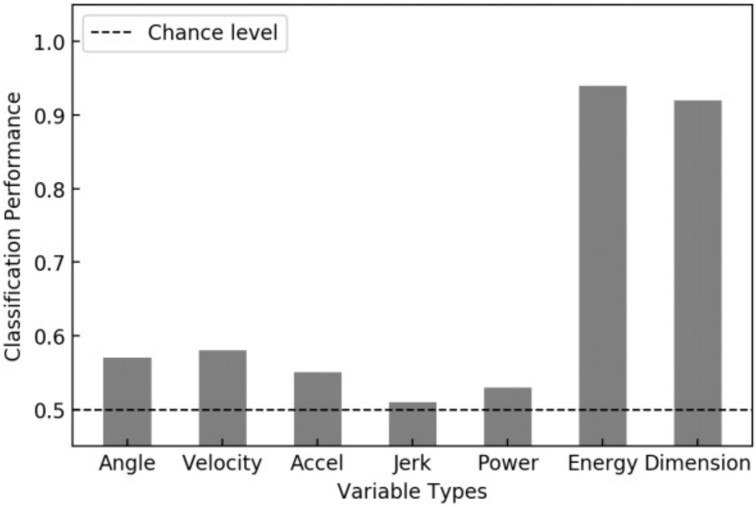

Figure 6: Results of classification tasks with several features.

an equally balanced number of samples from the two classes ($5000 - k + 1$ sample points for each class), the chance level for classification was 50%. The classification accuracy with angle (angular position), angular velocity, angular acceleration, angular jerk, and power spectrum was near or slightly above chance level. The accuracy with mechanical energy was approximately 95%, significantly higher than the chance level. This result is as expected: two classes of data generated using two distinct controllers that represent two distinct goals (i.e., goal angle G in equation 2.3) defined by mechanical energy. We treated this accuracy with mechanical energy as the best-possible reference accuracy in this classification task. Compared with this best-possible accuracy, the classification accuracy with pointwise dimension was approximately 92%, which was comparable. Note that unlike mechanical energy, which requires prior knowledge of the demonstrator's physical properties, pointwise dimension was computable using only a time series of angles formed by the pendulum that were observable to a naive imitator. This result suggests that pointwise dimension is a potentially useful feature for recognizing the intentions (controllers) behind observed movements that can be ascertained using only observable data.

To determine why both mechanical energy and pointwise dimension were effective for discriminating the latent underlying controllers, we further visualized how those features characterize the observable trajectories of the pendulums. Figure 7 shows how the position $y_t = \cos\theta_t$ of the manipulated pendulum, its mechanical energy, and pointwise dimension are correlated within the training data X_F for p_F (or X_A for p_A). For visibility, the

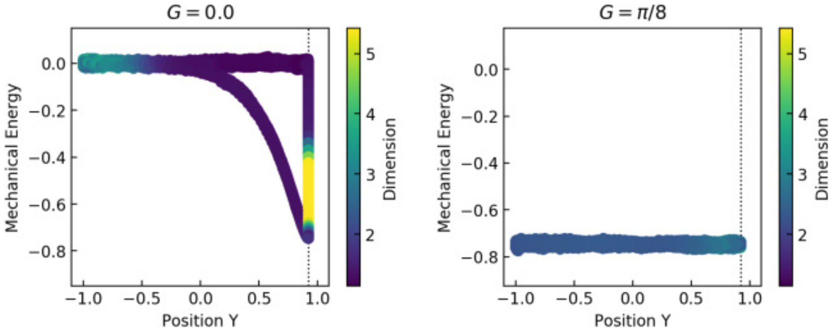

Figure 7: Dynamics of the pendulums in the plane of the y position and mechanical energy. The color of the data points indicates the pointwise dimension. The dashed lines indicate the bound $\cos\theta_{\text{min}}$.

value of the pointwise dimension $d(x)$ for each data point $x \in X_F$ (or X_A) in the figure was spatially averaged over the 100 nearest neighbors of x on the plane of the y position and mechanical energy. For the goal-achieved controller, in Figure 7 (right), the mechanical energy was mostly maintained at $E(G, 0)$ with $G = \theta_{\text{min}}$, based on the design of the energy-based controller. The estimated pointwise dimension was also mostly constant over time. In contrast, for the goal-failed controller, in Figure 7 (left), the mechanical energy and pointwise dimension both dramatically changed over time. According to the design of the controller, the mechanical energy was maintained at about $E(G, 0)$ with $G = 0$ for the feasible angle space $\theta_t > \theta_{\text{min}}$, that is, $y_t < \cos\theta_{\text{min}} \approx 0.923$, but decreased when the pendulum touched the bounds, remaining so until it started to leave the bounds with $E(\theta_{\text{min}}, 0)$. We think this difference in mechanical energy could contribute to high classification accuracy based on the mechanical energy.

Additionally, Figure 8 shows the direct relationship between the y position and the pointwise dimension when spatially averaged in the manner above. As shown in Figures 7 (left) and 8, the pointwise dimension first decreased around the time the pendulum touched the bounds and later increased when the pendulum started to leave the bounds, that is, the energy-based controller regained control of the pendulum. This tendency for a dramatic change in the pointwise dimension near the bounds can be observed in almost all other simulation runs. Thus, we hypothesized that the pointwise dimension characterizes such participation of the additional number of control variables, at least near the bounds of the infeasible angle space. Again, the great difference in pointwise dimension illustrated in these figures is expected to contribute to the high classification accuracy based on the pointwise dimension.

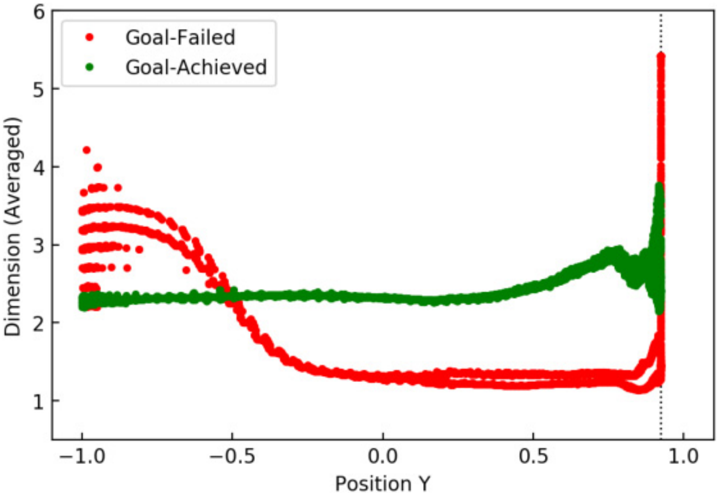

Figure 8: Dynamics of the pointwise dimension (averaged) of the pendulums as a function of their y position.

4 Simulation II: Action Completion Task

One of the key observations in the experiment conducted by WT is that the children could perform an action to achieve the demonstrator's "goal" by simply observing their incomplete action. Because the children did not observe the complete action in the experiment, they needed to identify the putative complete action by extrapolating the observed incomplete action. To explore the mechanism of the action completion task, we asked, how does the imitator observing the goal-failed demonstration produce an action that achieves the unobserved goal? As pointwise dimension was found to be reasonably characteristic of the intentions behind observed movements in simulation I, we examined an extended use of the pointwise dimension for the action completion task in this simulation.

In the action completion task, exact identification of the intention is not necessarily required or beneficial because the imitator (e.g., child) does not necessarily have the same body as the demonstrator (e.g., adult), and the motor controllers of different bodies required to meet the same goal may generally differ. Thus, in the action completion task, the imitator needs to identify two actions that have similar goals but may have different physical properties and latent motor control.

4.1 Action Completion Model. Based on the requirement described above, here we propose using the similarity in the dynamic transition patterns in the DoF of the two action-generating systems. Specifically, we hypothesize that the imitator observes an action and extracts the dynamics of the DoFs, defined by pointwise dimension, from the action as estimated for the recognition task in simulation I. Next, the imitator (mentally) simulates a movement by a given pendulum for each set of candidate controllers. Then the imitator performs an action by choosing the controller that can generate the action that is most similar to the demonstrated action. In this way, this action completion model uses a similarity in DoF dynamics rather than a similarity in apparent features such as angle and angular velocity patterns, which were found to be less characteristic of intentional differences in actions.

Specifically, we suppose that the imitator is exposed to one time series of angles generated in the goal-failed condition (see Figure 3C), which is suboptimal for the swing-up-no-hit task. We assume that the imitator performs an action by choosing a controller (see equation 2.3) with goal angle G as the parameter. In one condition, the other physical parameters, mass m and length l, of the pendulum are fixed at $(m, l) = (1, 1)$, the same values used by the demonstrator. In the two other pendulum conditions, the imitator uses either $(m, l) = (4, 1)$ or $(m, l) = (1, 2)$, which differ in mass or length to that used by the demonstrator. These three conditions are designed to investigate the robustness of the action completion model compared to differences in the physical features of the imitator's and demonstrator's pendulums. Given the goal-failed action, the action completion task of the imitator is to choose the controller f_G with the goal angle G that will most likely generate a movement similar to the demonstrated movement in terms of DoF dynamics. We let a set of controllers with the goal angles $0, 0.05, 0.1, \ldots, 0.9$ represent the imitator's options. The similarity in DoF dynamics of actions is described in the next section.

4.2 Similarity in DoF Dynamics. In this study, the DoF dynamics of a system are defined by the temporal change in the pointwise dimension estimated in the time series generated by the system. Specifically, a pointwise dimension estimator was constructed for a given demonstrated movement using the method proposed by Hidaka and Kashyap (2013), and used to estimate a series of pointwise dimensions for each of the demonstrated and candidate movements. The constructed pointwise dimension estimator is a mixture of multiple Weibull-gamma distributions, where each probability distribution P_i corresponds to a particular pointwise dimension and assigns for each point x_t in a trajectory the probability $P_i(x_t)$ that the point belongs to the ith distribution.

In our action completion model, the imitator is expected to perform an action controlled by the goal angle G that maximizes the log-likelihood function (defined by equation 4.4), which indicates some similarity in DoF dynamics between the demonstrated and simulated trajectory. Specifically, the log-likelihood is defined and maximized using the following steps (illustrated in Figure 9):

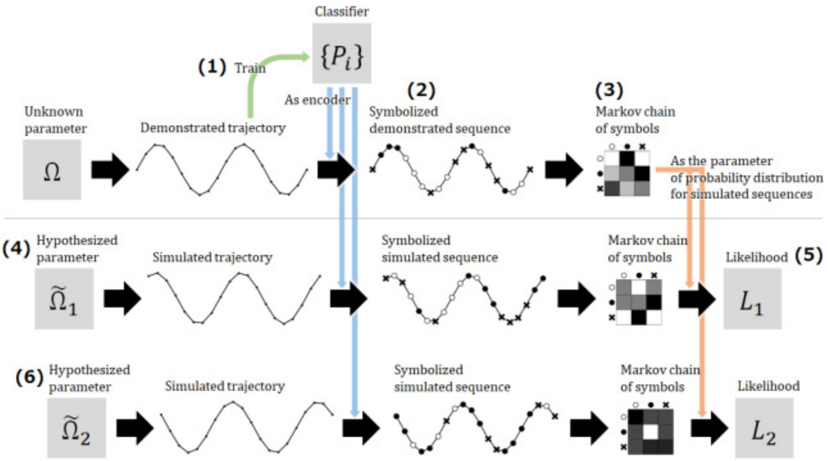

Figure 9: Our computational framework used for maximum likelihood inference for unknown parameters of the demonstrator behind the observed demonstrated trajectory. See text for details.

1. Given the demonstrated trajectory $\{x_t\}_{t\in T}$ as primary data, construct the pointwise dimension estimator/classifier $\{P_i\}_{i\in\{1,\ldots,k\}}$ for a reconstructed attractor by the method of time-delay coordinates with sufficiently high embedding dimension. The number of Weibull-gamma distributions k is chosen based on the Akaike information criterion (Akaike, 1974).
2. The demonstrated trajectory $\{x_t\}_t$ is transformed into a state sequence $\{s_t\}_t$, where each symbol $s_t \in \{1,\ldots,k\}$ is the index of the most likely distribution:

$$s_t = \arg\max_i P_i(x_t). \tag{4.1}$$

3. Denote $n^s_{i,j}$ as the number of transitions from state $s_t = i$ to state $s_{t+1} = j$ in $\{s_t\}_t$. Then the state transition joint probability matrix $Q \in \mathbb{R}^{k\times k}$ for all pairs of states is defined by

$$Q_{ij} = \frac{n^s_{i,j}}{\sum_{i'=1}^{k}\sum_{j'=1}^{k} n^s_{i',j'}}. \tag{4.2}$$

4. Given a candidate controller including its parameters (e.g., goal angle G), a simulated trajectory $\{y_t\}_{t\in T}$ is generated and transformed into another state sequence $\{u_t\}_t = \{\arg\max_i P_i(y_t)\}_t$. To calculate the probability $P_i(y_t)$, the simulated trajectory was first transformed by the method of time-delay coordinates with the same embedding

dimension, and the spatial neighborhood of sample point y_t was calculated within the feature space spanned by $\{y_t\}_{t\in T}$ itself. Then the state transition frequency matrix $H \in \mathbb{N}_0^{k\times k}$ is defined by

$$H_{ij} = n^u_{i,j}, \tag{4.3}$$

where $n^u_{i,j}$ is the number of transitions from state i to state j in $\{u_t\}_t$.

5. The likelihood of the candidate controller is defined by the multinomial distribution of a transition frequency H generated by the unknown controller showing the transition probability Q of the pointwise dimension estimator. Specifically, the log-likelihood function of the candidate controller with the goal angle parameter G is given by

$$\log L(G) = C_H + \sum_{i=1}^{k}\sum_{j=1}^{k} H_{ij} \log Q_{ij}, \tag{4.4}$$

where C_H is the multinomial coefficient for the multinomial distribution. To obtain comparable log-likelihoods for trajectories of unequal length, the log-likelihoods should be normalized by the numbers of state transitions, that is, $\sum_{i,j} H_{i,j}$.
6. Repeat steps 4 and 5 for other candidate controllers with different parameters.

This method is designed to abstract away differences in the absolute value of the pointwise dimension at each step between the two systems and compute similarities in the temporal change in the relative degrees of freedom. In our simulations below, the imitator is expected to generate an action trajectory controlled by the goal angle G that maximizes the log-likelihood function, equation 4.4.

We have two remarks on the procedure described above. First, in step 4, the spatial neighborhood of the sample point was calculated so as to abstract away differences in the physical properties of the pendulum, such as the absolute value of the positional data, such as differences in the pendulum length between the imitator and demonstrator. In the previous section, the neighborhood was calculated within the space spanned by the training data for the dimension estimator. However, this is valid only for the recognition task, in which we assumed the same physical properties for the two demonstrators. Second, in this method, we constructed Q from a demonstrated and H from a simulated trajectory. In terms of imitation or mimicry in biology, the demonstrator is actually the "model" or "exemplar" to be imitated by the imitator or "mimic"; hence, the mimic H can be seen as a realization of the model Q. However, the opposite view is also acceptable: the imitator forms a hypothesis Q' and tests it using a realization H' by the demonstrator. The same result will be obtained using either

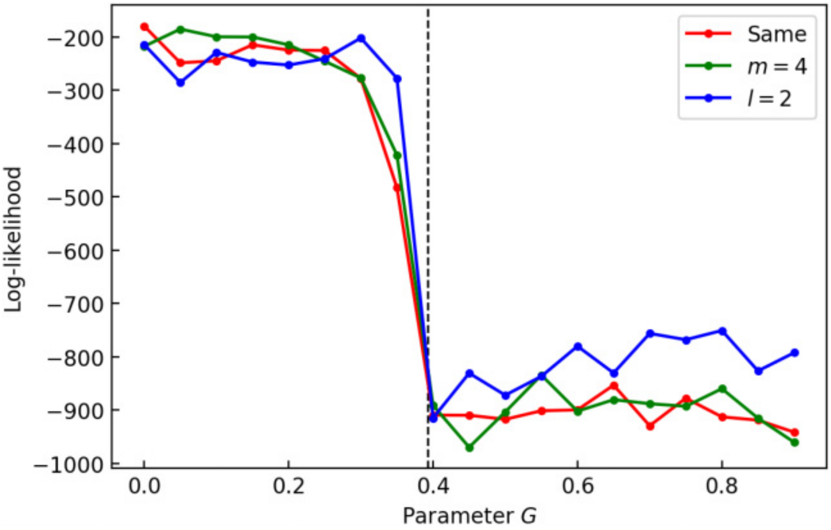

Figure 10: Results of the action completion task. The log-likelihood was computed using the action completion model. For the same-pendulum condition, the ground truth is $G = 0$. The vertical dashed line is the boundary of the feasible angle $\theta_{\min} = \pi/8$. For the different-pendulum conditions, candidate movements were generated using an imitator's pendulum of either $(m, l) = (4, 1)$ or $(1, 2)$.

view because similarity is a symmetric measure between the demonstrator and imitator, or the model and mimic. In the proposed method, we take the former view for the practical reason that the dimension estimation is computationally expensive and the estimation is required only once. This is in contrast to the latter view, which requires as many estimations as there are hypotheses.

4.3 Results of Action Completion. In this simulation, we assumed that the demonstrator had a controller with goal angle $G = 0$ but failed to meet the goal due to the infeasible region with its boundary at $\theta_{\min} \approx 0.39$. Thus, we set the ground truth at an estimated $G = 0$, as the latent goal angle of the controller used by the goal-failed demonstrator was $G = 0$. In the action completion task, the imitator is expected to generate an action that matches the latent intention of the demonstrator. The imitator employs the action completion model described above to produce the action most likely to display DoF dynamics similar to that of the demonstrator's.

For each goal angle G in the candidate set $\{0, 0.05, \ldots, 0.9\}$, we computed the log-likelihood function, equation 4.4, based on the similarity in DoF dynamics between demonstrated and simulated movements (see Figure 10).

For each G, the figure shows the average of 20 log-likelihood values over sampled simulated actions with different initial values. In the action completion model, a controller with some G with higher log-likelihood is more likely to be chosen as the action produced by the imitator. Figure 10 (red points) shows the log-likelihood of goal angle G in the same-pendulum condition, that is, when the demonstrator and imitator control the same pendulum with the same physical parameters ($m = 1, l = 1$). Although the log-likelihood was not the highest at $G = 0$, it was generally higher for one range of angles $0 \leq G < \theta_{\min}$ than others $G \geq \theta_{\min}$. When we separated these angles into two groups at the boundary angle $\theta_{\min}$, we found the log-likelihood values on average were significantly different ($t(376) = 39.76$, $p < 0.01$). Thus, based on the similarity in DoF dynamics, these results suggest that the imitator can generally differentiate between two latent types of candidate action, which correspond to the difference between the swing-up and swing-up-no-hit task. In other words, the imitator can estimate that the demonstrator's latent intention is to move up beyond the infeasible boundary rather than to stop at the boundary as suggested in direct observation.

To what extent does this action completion model depend on the sameness of physical parameters of the demonstrator's and imitator's pendulums? To examine the robustness of this action completion model against deviations from the identical physical setting ($m = 1$, $l = 1$) between the demonstrator and imitator, we analyzed the same action completion task when the imitator controls pendulums with different physical parameters ($m = 4$ and $l = 1$, and, $m = 1$ and $l = 2$). In both cases, we obtained essentially the same results (green and blue points in Figure 10) as those in the same-pendulum condition (red points in Figure 10). The two groups of log-likelihood values were on average both significantly different ($t(375) = 40.93$, $p < 0.01$ for the condition with $m = 4$; $t(374) = 34.88$, $p < 0.01$ for the condition with $l = 2$). That is, even with physically different pendulums, the imitator can differentiate between the two general types of intentions (i.e., swing-up versus swing-up-no-hit). Thus, by using DoF dynamics as an indicator of similarity in movements, the imitator can successfully abstract away differences in physical features of the two pendulums. Note that when using two physically different pendulums, the motor controller has no "ground truth" of goal inference or no actual way of producing a movement that is exactly the same as the demonstrator's. Thus, in these different-pendulum conditions, it would be difficult for a simple movement-matching strategy to reproduce some unseen action performed by the demonstrator.

The previous experiment in which $G = 0$ and bounds of the infeasible region $\theta_{\min} = \pm\pi/8 \approx 0.39$ was conducted to test the robustness of the action completion model against nonidentical latent body parameters, l and m. Because the bounds of the infeasible region can also be latent bodily parameters, we conducted a similar line of experiments with $G = 0$ and various hypothetical bounds $\theta_{\min}$ for the imitator's pendulum, or the "mental

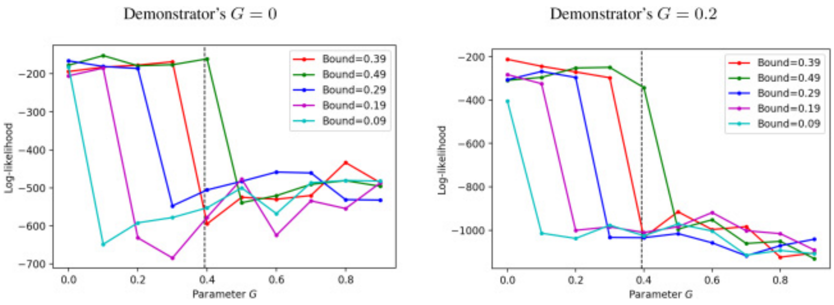

Figure 11: Results of the action completion task. The log-likelihood was computed using the action completion model. The ground truth is $G = 0$ for the chart on the left and $G = 0.2$ for the chart on the right. The vertical dashed line is the boundary of feasible angle range $\theta_{\min} = \pi/8$. Candidate movements were generated by the imitator's pendulum with various hypothetical bounds $\theta_{\min} = 0.09, 0.19, 0.29, \pi/8 \approx 0.39, 0.49$.

simulator." In these experiments, the imitator can also manipulate the infeasible bounds parameter $\theta_{\min}$, called the hypothetical bounds. Figure 11 (left) shows the goal inference results using pendulums with different bounds for the infeasible region $\theta_{\min} = 0.09, 0.19, 0.29, \pi/8 \approx 0.39, 0.49$. The ground truth parameter (of the demonstrator's pendulum) is $\theta_{\min} = \pi/8 \approx 0.39$. This experiment showed that the imitator inferred that some goal angle greater than her pendulum's hypothetical bounds is more likely, regardless of the pendulum's actual hypothetical bounds. Figure 11 (right) shows the results for the same line of experiments with $G = 0.2$. The results are similar to those in Figure 11 (left) despite the difference in goal G. Thus, our proposed method for goal inference can be performed without knowledge of the physical parameters l, m or the bounds of the infeasible region $\theta_{\min}$. However, our method cannot currently be used to identify the exact goal G, but simply to infer a range within which the goal G falls.

An additional point that should be examined is whether goal inference is possible from the observation of nonrepeated actions. The new goal inference task will be referred to as single-shot goal inference, in which the demonstrator does not repeat a similar (goal-failed/achieved) movement but provides only one swing of the pendulum such that it moves from the bottom to the top-most position and then to the bottom again. In this task, the imitator observed a batch of one-swing movements performed by the demonstrator aiming to obtain the same goal angle; the imitator subsequently inferred the latent goal angle. For training, 30 one-swing movements by the demonstrator were provided, and an embedding dimension of 20 was used. For inference, we restricted our imitator such that it likewise could produce only one-swing movements in his mental simulation.

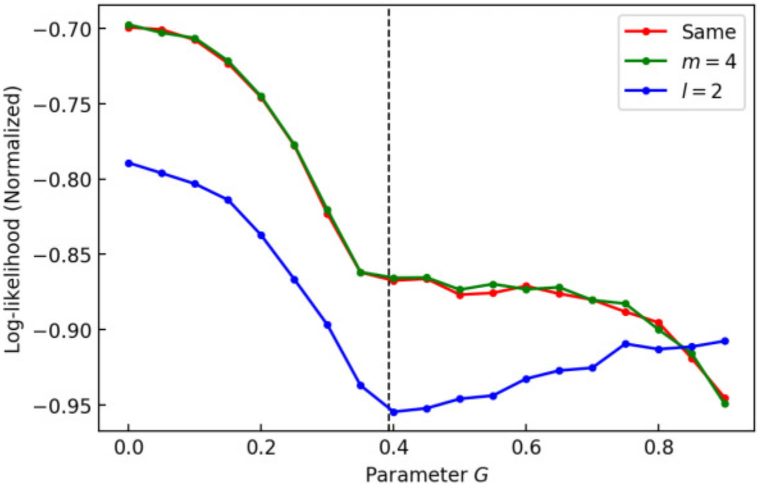

Figure 12: Results of the single-shot action completion task. The normalized log-likelihood was computed. See the Figure 10 caption for details.

A batch of one-swing movements, except for the final one, for training can be interpreted as a collection of the imitator's past experiences in the same or a similar situation as that of the demonstrator. Similar to Figure 10, Figure 12 shows the average of 20 normalized values of the log-likelihood of goal angle G in the same and different pendulum conditions. The results for nonrepeated actions in general show a similar tendency to our previous results for repeated actions. This result suggests the robustness of the proposed goal inference scheme against this type of data shortage.

4.4 Goal Inference via Inverse Reinforcement Learning. One of the standard techniques used to infer a goal from observed data on an action is inverse reinforcement learning (IRL) (Ng & Russell, 2000). In IRL, the reward function is inferred from a large number of state-transition sequences based on the assumption that those sequences were sampled from an unknown Markov decision process. We adopted one IRL algorithm (Ziebart et al., 2008) that has been frequently reported to be robust and efficient in the IRL literature. The basic idea of IRL is often represented as frequency matching: in general, IRL algorithms estimate a higher reward for a more frequently visited state. Because the pendulum swing-up task is quite common in reinforcement learning (Sutton & Barto, 1998; Doya, 1999), we adopted the simplest state-space discretization method, called tile coding. The state space $(\theta, \dot{\theta}) \in (-\pi, \pi] \times (-2\pi, 2\pi]$ is divided equally into 64 × 64 equally spaced tiles. Time series are sampled every three time steps.

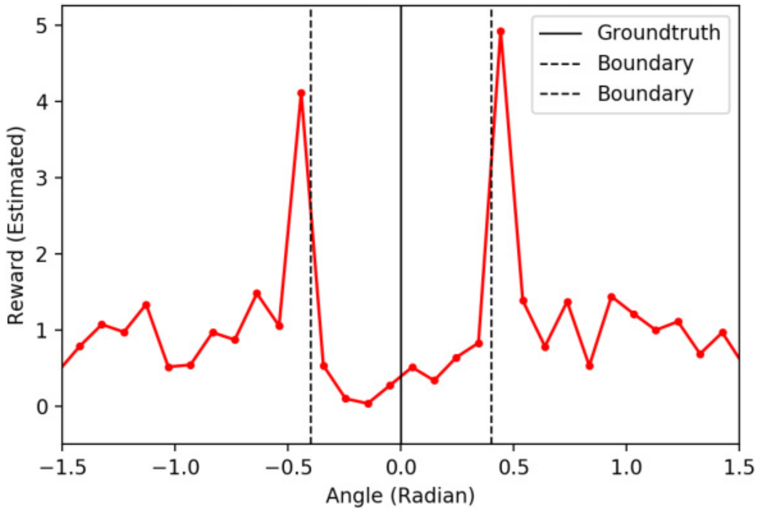

Figure 13: Reward function predicted by an IRL algorithm. The ground truth is $G = 0$. The vertical dashed lines are the boundaries of the feasible angle $\theta_{\text{min}} = \pi/8$.

Figure 13 shows the reward function estimated by the IRL algorithm for observed state sequences generated from a movement performed by the goal-failed demonstrator with goal angle $G = 0$. Since her pendulum is constrained by the infeasible region, the controller constantly hits the bounds of the infeasible region, and thus the most frequently visited states are those in which the pendulum stops at the bounds of the feasible region $(\theta, \dot{\theta}) \approx (\pm\pi/8, 0)$. As shown in Figure 13, the IRL algorithm estimated that the highest reward states were at the bounds of the feasible region $\theta = \pm\pi/8$, and the actual $G = 0$ was estimated as a low-reward state, which is unlikely to be the demonstrator's goal angle. This result is expected, as IRL generally works as a "frequency-matching" algorithm. Therefore, our findings suggest that a frequency-matching algorithm such as IRL works poorly when it is used to complete an unobserved action (with zero frequency) in the goal-failed situation.

5 Discussion

Inspired by the psychological experiment conducted by Warneken and Tomasello (2006), we designed a minimal simulation framework to account for the mechanism of action recognition and action completion. We showed that the simulated imitator can discriminate between goal-failed and goal-achieved actions, which have apparently similar movements but different intentions and goals (simulation I). Then we proposed an action completion model that can perform an action comparable to the optimal action for the swing-up task simply by observing the goal-failed action, which is suboptimal with the pendulum in the infeasible region (simulation II). Both recognition and completion can be the basis of goal inference from an unsuccessful demonstration.

In these two simulations, we used DoF dynamics in actions, or a type of abstraction of bodily movements using a dynamical invariant, as a feature of the underlying motor controllers. For this abstraction, the obtained DoF dynamics can effectively ignore apparent positional variation among observed movements while extracting the dynamical/mechanical characteristics behind the movements. Our simulations comparing action completion based on DoF dynamics versus the frequency of spatial/positional states (see Figure 10 versus Figure 13) suggested that our abstraction to DoF dynamics allowed the imitator to identify a range of controllers (with the parameter $G > \theta_{\min}$) including the optimal controller for the demonstrator's latent goal. Our additional simulations (see Figure 11) suggested that this abstraction may not allow the imitator to exactly identify the demonstrator's goal. We consider that this limitation of our proposed method is acceptable, as even we humans cannot infer another's hidden goal exactly but can rather identify the general direction of the demonstrator's intended goal. For example, consider the case in which you seeing a man kicking a closed door many times while both of his hands are full. You may think that he wants to open the door. But how can you infer exactly where he is heading after he goes through the door? It is difficult for you to determine this without any prior knowledge of his goal. In our simple pendulum simulations, the imitator successfully inferred that the demonstrator wanted to go beyond the infeasible bounds (the door) but could not exactly identify the demonstrator's goal angle (where he is heading). Given the theoretical results of this letter—that DoF dynamics are effective in both action recognition and action completion—we predict that this feature will also play a crucial role in understanding human action and imitation. This hypothesis will be tested in future work.

Finally, we add two remarks on why a dynamical invariant is effective for completing a goal-failed action compared with existing approaches. First, our approach using a dynamical invariant does not presume any kind of optimality of observed actions, whereas existing approaches, such as inverse optimal control (Wolpert, Doya, & Kawato, 2003) and IRL (Ng & Russell, 2000), do. This difference between the approaches is crucial, because the observed action, to be completed, *failed* in its original goal in our task.

Second, our approach using a dynamical invariant is likely to be useful for estimating the point-to-control underlying the goal-failed or goal-achieved actions. In general, bodily movements need to be more carefully controlled when the movement is at a state closer to the final goal. Consider reaching, for example. Finer control is needed near the point to be reached, more so than at the beginning of the action. This necessary control gain may be reflected by the observable granularity in the fluctuation of an action and can be quantified using a type of fractal dimension of trajectories.

As Bernstein (1996) pointed out, while the DoF of our bodies can be a source of flexibility in our movements, generally a system with a very large number of DoF is likely to be intractable. Therefore, organisms must reduce their body's DoF to be tractable (Bernstein, 1996). We predict that a reduction in DoF is crucial especially when accuracy of movement is required or when one is close to the task goal state. Thus, we speculate that a dynamic decrease or increase in DoF might inform the imitator about whether the current state of the observed system is near or far from the unknown task goal state. Specifically, the goal-failed actions we adopted in this letter can have their own characteristic dynamic pattern of DoF—for example, the different ways by which the pendulum touches the boundaries of the infeasible space. Therefore, our approach can successfully infer the hidden goal of an observed action, even if the observed action is suboptimal and goal-failed.

A review of IRL (Zhifei & Joo, 2012) proposed that learning from goal-failed or imperfect or incomplete demonstrations is a challenging new problem in related research fields. Our approach based on a dynamical system is expected to bring new insights to this class of problems.

The proposed model, at least in a minimal physical model such as a pendulum control task, is reasonably effective for an action completion task. Although we hypothesized that DoF can be a commonly effective feature for goal inference, the evidence for this claim (i.e., the results from our simulation models) is limited as we supposed that the simple pendulum was a physical body and that we had a sufficient amount of training data, among other assumptions. Given that the simple pendulum is a mechanical system with only one DoF, this assumption greatly simplifies the problem of imitation, which may require a large number of DoFs to control the body. In imitating systems with high DoF, an ill-posed problem, in which there are multiple different ways to achieve the same goal, is another fundamental problem that we have not addressed in this letter. Furthermore, we only quantitatively studied simple stage goal-directed actions with no need for explicit subgoals, as opposed to complex actions composed of multiple stages with the need for explicit subgoals. Whether DoF can be generally effective or how it can be effectively exploited for goal inference from complex actions of high DoF systems should be explored in the future. We expect to extend the current work to more complex action-generating systems in the future.

Acknowledgments

We thank the anonymous reviewer. This work was supported by JSPS KAKENHI grants JP 20H04994 and JP 16H05860.

References

Abbeel, P., & Ng, A. (2004). Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the 21th International Conference on Machine Learning* (pp. 1–8). PMLR. https://doi.org/10.1145/1015330.1015430

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Argall, B., Browning, B., & Veloso, M. (2007). Learning by demonstration with critique from a human teacher. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (pp. 57–64). New York ACM. https://doi.org/10.1145/1228716.1228725

Astrom, K. J., & Furuta, K. (2000). Swinging up a pendulum by energy control. *Automatica*, *36*(2), 287–295. https://doi.org/10.1016/S1474-6670(17)57951-3

Babes-Vroman, M., Marivate, V., Subramanian, K., & Littman, M. (2011). Apprenticeship learning about multiple intentions. In *Proceedings of the 28th International Conference on Machine Learning* (pp. 897–904).

Bernstein, N. A. (1996). *Dexterity and its development*. London: Psychology Press.

Breazeal, C., & Scassellati, B. (2002). Robots that imitate humans. *Trends in Cognitive Sciences*, *6*(11), 481–487. https://doi.org/10.1016/S1364-6613(02)02016-8

Cutler, C. D. (1993). A review of the theory and estimation of fractal dimension. In H. Tong (Ed.), *Dimension estimation and models* (pp. 1–107). Singapore: World Scientific.

Doya, K. (1999). Reinforcement learning in continuous time and space. *Neural Computation*, *12*, 243–269. https://doi.org/10.1162/089976600300015961

Grollman, D. H., & Billard, A. (2011). Donut as I do: Learning from failed demonstrations. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 3804–3809). Piscataway, NJ: IEEE. https://doi.org/10.1109/ICRA.2011.5979757

Grollman, D. H., & Billard, A. (2012). Robot learning from failed demonstrations. *International Journal of Social Robotics*, *4*, 331–342. https://doi.org/10.1007/s12369-012-0161-z

Hidaka, S., & Kashyap, N. (2013). *On the estimation of pointwise dimension*. arXiv:1312.2298.

Ho, M. K., Littman, M., MacGlashan, J., Cushman, F., & Austerweil, J. L. (2016). Showing versus doing: Teaching by demonstration. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems*, *29* (pp. 3027–3035). Red Hook, NY: Curran.

Hogan, N. (1984). An organizing principle for a class of voluntary movements. *Journal of Neuroscience*, *4*(11), 2745–2754. https://doi.org/10.1523/JNEUROSCI.04-11-02745.1984, PubMed: 6502203

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838–850. https://doi.org/10.1023/A:1013251112392

Ng, A., & Russell, S. J. (2000). Algorithms for inverse reinforcement learning. In *Proceedings of the 17th International Conference on Machine Learning* (pp. 663–670).

Schaal, S. (1997). Learning from demonstration. In M. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems*, *9* (pp. 1040–1046). Cambridge: MIT Press.

Schaal, S. (1999). Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, *3*(6), pp. 233–242. https://doi.org/10.1016/s1364-6613(99)01327-3, PubMed: 10354577

Shiarlis, K., Messias, J., & Whiteson, S. (2016). Inverse reinforcement learning from failure. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems* (pp. 1060–1068). New York: ACM.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tomasello, M. (2001). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press. https://doi.org/10.2307/j.ctvjsf4jc

Torii, T., & Hidaka, S. (2017). Toward a mechanistic account for imitation learning: An analysis of pendulum swing-up. *New Frontiers in Artificial Intelligence*, *10247*, 327–343. https://doi.org/10.1007/978-3-319-61572-1_22

Uno, Y., Kawato, M., & Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement: minimum torque-change model. *Biological Cybernetics*, *61*(2), 89–101. https://doi.org/10.1007/BF00204593, PubMed: 2742921

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311*, 1301–1303. https://doi.org/10.1126/science.1121448, PubMed: 16513986

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 593-602.

Young, L.-S. (1982). Dimension, entropy, and Lyapunov exponents. *Ergodic Theory and Dynamical Systems*, *2*(1), 109–124. https://doi.org/10.1017/S0143385700009615

Zhifei, S., & Joo, E. M. (2012). A review of inverse reinforcement learning: theory and recent advances. In *Proceedings of the IEEE World Congress on Computational Intelligence* (pp. 1–8). Piscataway, NJ: IEEE. https://doi.org/10.1109/CEC.2012.6256507

Ziebart, B. D., Maas, A., Bagnell, J. A., & Dey, A. K. (2008). Maximum entropy inverse reinforcement learning. In *Proceedings of the 23rd National Conference on Artificial Intelligence* (vol. 3, pp. 1433–1438). Palo Alto, CA: AAAI Press.

Received October 23, 2020; accepted June 4, 2021.