How can we help others?: a computational account for action completion

Takuma Torii (tak.torii@jaist.ac.jp) Shohei Hidaka (shhidaka@jaist.ac.jp)
Japan Advanced Institute of Science and Technology
1-1 Nomi, Ishikawa, Japan

Abstract

To help others, we need to infer one's goal and intention and make an action which complements one's action yet to meet the underlying goal. In this study, we consider the computational mechanism how a person can infer the other's intention and goal from his or her action, which is not completed or fails to meet the goal. As a minimal motor control task toward a goal, we analyzed single-link pendulum control tasks and its variation. By analyzing two types of pendulum control tasks, we show that a sort of fractal dimension of movements is characteristic of the difference in the underlying motor controllers. Further, using the fractal dimension as a criterion of similarity between movements, we show that the simulated pendulum controller can make an action toward the goal, toward which other's incomplete action was made, but was not observable in behavior due to its failure.

Keywords: imitation; intention; action; motor control; dynamical system; fractal dimension;

Introduction: Imitation of Action

As a way of social learning from others, children imitate their parents' movements in an early developmental period. Imitation, as a behavioral basis for understanding other's goal and intention, is thought of a mechanism to preserve social and cultural knowledge. Thus, from the perspective of cultural evolution, it plays a key role as a "latchet" preventing human culture from stepping back (Tomasello, 2001).

In a typical imitation, a demonstrator (e.g., parent) shows an imitator (e.g., child) an *action* with an *intention*. In this paper, by "intention" we mean either a motor plan or a motor control toward a certain goal, that gives a series of choices on each step in order to achieve a given goal, and by "action" we mean a movement with an intention to achieve a certain goal (Bernstein, 1996).

In this study, we consider a certain type of imitation learning, where the imitator does not know the demonstrator's goal and intention behind its action. Given this situation we pose, the imitation learning requires to solve the two major classes of problems: *identification of action features* with which two actions with different intentions can be discriminated, and *action completion* which extrapolates an incomplete part of other's action which fails to meet its goal, and make an action to meet the goal. The first problem, identification of features, requires features correlated to the intentional difference (functional difference in motor control) behind an action, rather than features which just describes apparent movements. Inference of an intention or a goal behind an action is, however, an ill-posed problem in general: a pair of two similar movements can be produced by two quite different intentions (motor plans) or by two different goals. The second problem, action completion, needs the identification of feature and to apply it to make an action to meet an estimated goal for an observed part of an incomplete action which the demonstrator intended to finish but failed.

In this study, we aim to address these two questions by a form of numerical study on a task to control a physical object – a single-stick pendulum. We suppose that this simple control task is minimally sufficient to capture the essential aspect of goal imitation: how one can recognize the intention (motor control) behind a given action, and how the one reproduces it.

Although the control task of a pendulum may be viewed overly simplified in its structural complexity compared to the actual human body, we view this task has essentially similar characteristics with the experimental task reported by Warneken & Tomasello (2006). In their experiment, 18 months olds were exposed to an adult (experimenter)'s goal-failed behavior, and they investigated whether those children could infer the adult's goal, which was not demonstrated there, and help to complete it. They have suggested children of this age can infer others' goals and complete the actions.

In principle, the child in their experiment is required (1) to recognize the failed goal and intention and (2) to make an action by controlling his/her own body to meet the goal. The task (1) and (2) are called *recognition* and *completion* task for goal imitation, respectively. In what follows, we illustrate how our simulation framework captures the goal imitation behavior, and then report two simulation studies for recognition and completion task.

Simulation Design

Rationale: Abstracting Warneken & Tomasello

Here we briefly introduce the experiment of Warneken & Tomasello (2006) (WT in short hereafter). WT have investigated whether children can infer the demonstrator' goal and the intention behind the behavior. In a situation, in the experimental condition, called out-of-reach situation, a demonstrator dropped a marker on the floor accidentally and could not reach for it, whereas in the control condition he intentionally dropped a marker on the floor. The former condition implicitly calls for child's help for the demonstrator to finish unsuccessful intention, namely to pick up the marker, but the latter does not. The pair of experimental and control condition was designed so that the two of the demonstrator's apparent bodily movements are similar (e.g., both dropped a marker), whereas the underlying intentions behind the actions were quite different. WT showed that the children showed helping behaviors more frequently in the experimental condition than the control condition.

In this study, we design a simulation framework so that it can capture the essence of WT's experimental design in a

minimal form. Specifically, we employed a single-link pendulum as a simplified human body. Each of the imitator (i.e., the hypothetical child) and the demonstrator is supposed to control a pendulum to make an action (i.e., a goal-directed movement). The *goal* of the demonstrator is set to keep the pendulum at the topmost position (opposite to gravity) as much as possible subject to a given "bodily constraint", i.e., a given set of physical parameters of the pendulum (mass, length, and so on). The *intention* of the demonstrator is its controller of the pendulum, which is angle acceleration (force) as a function of angle and angular velocity of the pendulum. An *action* of the demonstrator is a movement of the pendulum, represented either by the (x, y) coordinate or a vector of angle and angular velocity, which is generated by an initial condition and the controller of the demonstrator.

We think that the essential difference between the experimental (goal-failed) and the control (goal-achieved) condition in WT is captured by the degree of *optimality* of intention or action for a given goal. Suppose there are controllers A and B, which are optimal for different goal G_A and G_B, respectively. If the demonstrator uses control A for goal G_A, its generated movement would be optimal and treated as a "successful" action. While, if the demonstrator uses B for goal G_A of A, its generated movement would be sub-optimal and treated as a "failed" action. The former case is an analog to the control condition in WT and the latter is to the experimental condition.

Accordingly, we design two different tasks (combination of a goal and constraint) for demonstrators. (A) The *swing-up task* has the goal to keep the pendulum being as close as possible to the top without any obstruct (Figure 1A). The goal is quantified by the reward function of the angle θ defined by $r(\theta) = \cos\theta$, where the top position with $\theta = 0$. (B) The *swing-up-no-hit* task constrains the pendulum from moving to a certain region in angle (*infeasible region*: the black region shown in Figure 1B), and the goal is to keep the pendulum being as close as possible to the top unless it hits the infeasible region. In the task B, the demonstrator will be given the least reward when the pendulum is at the infeasible region (including its boundary); otherwise, the demonstrator is given reward as a function of angle ($r(\theta) = \cos\theta$; the least value is $r(\pi) = -1$) at each time step. The degree of optimality (match/mismatch between the intention and the movement) is defined by the cumulative reward function of a given movement over time relative to its maximum.

In the *goal-failed* condition of our simulation, that is the analog to the experimental condition of WT, the demonstrator works on the task B (with the infeasible region) by controlling the pendulum with the controller optimal for the task A (Figure 1C). In the *goal-achieved* condition, that is the analog to the control condition, the demonstrator work on the task B by the controlling it with the controller optimal for the task B (Figure 1D). Obviously, the demonstrator in the goal-failed condition (Figure 1C), but not the one in the goal-achieved one (Figure 1D), shows a movement sub-optimal for the task B, which does not match the "intended" movement being optimal for the task A.

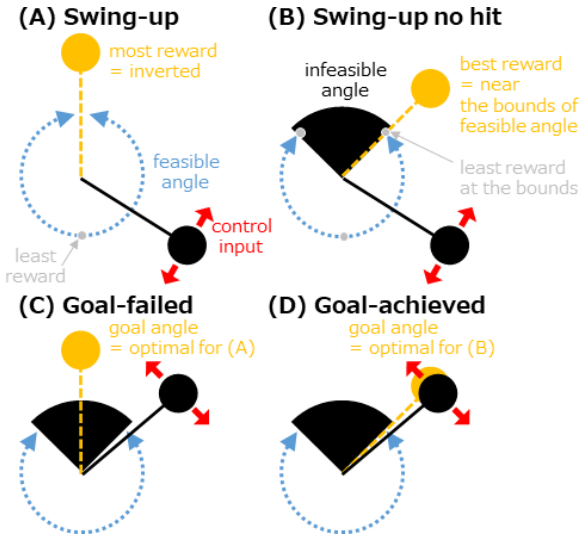

Figure 1: Simulation design analogous to experimental tasks in Warneken & Tomasello (2006). (A) The swing-up task. The most rewarding angle is the topmost ($\theta = 0$) and the least is the bottom ($\theta = \pi$). (B) The swing-up no-hit task. There are three least rewarding angles: the bottom ($\theta = \pi$) and the bounds of the infeasible region (black: $\theta = \pm\pi/8$). The best rewarding angle is somewhere closer to the top within feasible region. It is optimal to keep on swinging without touching the bounds. (C) Goal-failed demonstration: working on the task B with the control optimal for the task A. (D) Goal-achieved demonstration: working on the task B with the control optimal for the task B.

The imitator, in turn, observes these different actions in the two conditions. Here we provide two analyses for recognition and completion of the actions, which are respectively analog to the child's sub-tasks, recognizing the difference between the actions observed in the experimental and control condition, and making an action to complete the intended movement to help the unsuccessful demonstrator.

The Pendulum Control

A mathematical model of simple pendulums is composed of a link of length $l = 1$ and point mass $m = 1$ at the tip of the link. A state of this pendulum is determined by the angle θ of the link and angular velocity $\dot{\theta}$. The angle is relative to the inverted position $\theta = 0$. The equation of motion is given by

$$ml^2\,\ddot{\theta} - mgl\,\sin\theta = u + \varepsilon \qquad (1)$$

where $g = 9.8$ is the gravity constant, u is control input (torque) from a controller f, and $\varepsilon \sim N(0, \sigma)$ is intrinsic noise, a normally distributed random variable.

The pendulum swing-up task is a classic in feedback control theory (Doya, 1999), originally to design a controller that can swing the pendulum up and hold it about the inverted position given by $\theta = 0$. A controller for this task is defined by the function $u = f(\theta, \dot{\theta})$, which outputs torque u for a given state $(\theta, \dot{\theta})$. The goal of a demonstrator is to choose f so as to maximize $\sum_t r(\theta_t)$ (see the section later for details).

The initial position of the pendulum is set $\dot{\theta} = 0$ and a uniformly random value for $\theta = \pm[\pi/8, \pi)$.

Energy-based Swing-up Controller

The simple pendulum is characterized well by the mechanical energy that is the sum of the kinetic and potential energy:

$$E(\theta, \dot{\theta}) = \frac{1}{2} ml^2\dot{\theta}^2 + mgl\,(\cos\theta - 1). \qquad (2)$$

In the pendulum swing-up task, its goal state holding at the inverted position $\theta = 0$ and $\dot{\theta} = 0$ corresponds with $E(0,0) = 0$.Without any control input ($u = 0$) and noise ($\sigma = 0$), the simple pendulum preserves the mechanical energy over time. Thus, the goal of the swing-up task was met by controlling the mechanical energy be zero ($E(\theta, \dot{\theta}) = 0$). Employing this observation, Astrom & Furuta (2000) proposed the energy-based controller

$$f(\theta, \dot{\theta}) = -\left(E(\theta, \dot{\theta}) - E(G, 0)\right)\dot{\theta}, \qquad (3)$$

with which one can control the current energy $E(\theta, \dot{\theta})$ to be closer to the targeted energy $E(G, 0)$ with the goal angle G.

Goal-achieved and Goal-failed Action

For each of the demonstrator in the swing-up and swing-up-no-hit task, different reward function $r(\theta)$ is assumed. In the swing-up task, the reward function is $r(\theta) = \cos\theta$, that indicates the top-most position $\theta = 0$ is the most rewarding position for the pendulum. In the swing-up-no-hit task with the infeasible bound at the angle $\theta_{\min}$, the reward function is $r(\theta) = \cos(\pi) = -1$ if $\theta = \theta_{\min}$ and otherwise $r(\theta) = \cos\theta$. The optimal controllers are different for the two tasks with different reward functions. The optimal controller for the swing-up and swing-up-no-hit with the infeasible bound at the angle $\theta_{\min} = \pi/8$ are respectively the controller (Equation 3) with the goal angle $G = 0$ and $G = \theta_{\min}$.

Controlling the pendulum with Equation 3 with some goal angle $G \neq 0$ in the swing-up task, it results in swinging within the range $\theta > G$. In the swing-up-no-hit task, the pendulum is constrained within the feasible angle $\theta \in \pm(\theta_{\min}, \pi]$ due to the absorbing bound.

Figure 2 shows typical actions (time series of angles) made by the goal-failed demonstrator with the goal angle $G = 0$ (top) and goal-achieved one with $G = \theta_{\min}$ (bottom) working on the swing-up-no-hit (Figure 1C and D). Those movements look similar in angle dynamics, which simulates the movement similarity in WT (e.g., dropping a marker accidentally and intentionally), but their mechanical energy, a direct indicator of their controller, may show some difference between the two actions.

Feature for Intentional Difference

According to our definition of the goal-failed and goal-achieved condition, the intention behind movements, optimal for the swing-up task, mismatches the swing-up-no-hit task (Figure 1C). Not only in this particular case, many other "failed" actions, including WT's, are supposed to be with this type of mismatch between some originally intended task and the actually performing task. One of critical feature common

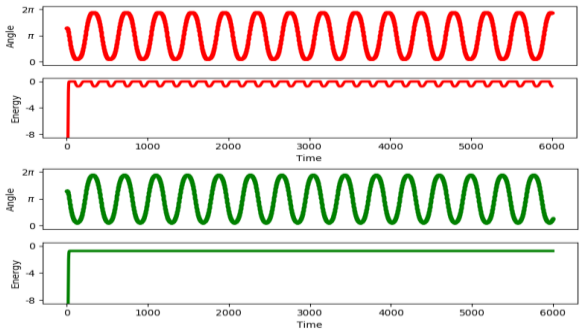

Figure 2: Representative time series of angle and mechanical energy made by the goal-failed demonstrator (top two panels) with $G = 0$, and the goal-achieved demonstrator (bottom two panels) with $G = \theta_{\min}$, in the swing-up-no-hit task

in this type is that the task to work has an additional obstacle which is not expected in the original task.

Beyond specific differences across different tasks, we hypothesize that this type of failures may be characterized by the existence of some additional factor complicating the originally intended task. In the dropping a marker accidentally in WT, the demonstrator was supposed not to ready for the situation that requires him to pick the dropped marker, and the accidental dropping gives the additional complexity to the originally intended task, to carry the marker to somewhere (without dropping it). So is the goal-failed condition in our pendulum simulation: the additional obstacle, the limitation in the feasible angle, causes the sub-optimality of the original motor control in this unexpected new task.

Accordingly, we suppose that this additional factor of complexity could be quantified by the degrees-of-freedom (DoF) of a given system. In the accidental dropping case again, in order to succeed the unintended task, the demonstrator needs to have some DoF to choose whether he flexibly changes the motor plan to solve the new situation (unexpected dropping). Similarly, the pendulum controller equipped with an additional DoF resetting the new target angle may flexibly tune his motor control to be near optimal for a given new situation. That means, in both two cases above and perhaps more generally, the unintended new task introduces some additional complexity or DoF to be solved by the demonstrator.

In this paper, we specifically employ a sort of fractal dimension, called pointwise dimension (see Section "Pointwise Dimensions"), of the actions as an indicator of the DoF of the underlying controller, and test whether it is characteristic of the intentional difference underlying the actions. In the following two sections, we examined our hypothesis by the two simulations for (1) recognition task and (2) completion task from the imitator's perspective.

Simulation I: Recognition Task

In Simulation I, we investigate whether the imitator can tell the two different intention underlying the actions in the goal-failed and goal-achieved condition. The goal of this

simulation is to analyze and identify which feature is more characteristic of the latent intention of actions.

Specifically, we listed angle, angular velocity, power spectrum, mechanical energy, and pointwise dimension of movements for this analysis. The angle, angular velocity, and power spectrum are standardly employed features of movements in the literature. It is also natural for our simulation, as the motor control is a function of angle and angular velocity, and the generated movement is periodic. The mechanical energy is indeed the very term defining the motor control (see Equation 3), and thus we expect that the mechanical energy would be the best possible feature in theory to characterize the intention (motor control). It is, however, that a naive imitator such as a child may not be able to directly access the mechanical energy, as it needs knowledge of the physical parameter of the pendulum (i.e., mass m and length l in Equation 1). Thus, the mechanical energy is treated as an indicator for the best-possible recognition performance in our analysis.

Lastly, pointwise dimension is a feature indicating the latent DoF of the underlying dynamical system, and we hypothesize that it is an indicator of task complexity and would be characteristic of intentional difference between movements in the goal-failed and the goal-achieved condition. Our hypothesis predicts that pointwise dimension is a characteristic feature as good as mechanical energy in the classification of the movements with different intentions.

Pointwise Dimensions

To characterize complexity in the demonstrator's movements, we analyze the attractor dimension of the movements treating it as a dynamical system. Specifically, we exploited a sort of fractal dimension called pointwise dimension for the classification analysis. The pointwise dimension is a type of dimension, which is defined for a small open set or measure on it including a point in a given set (see Cutler, 1993; Young, 1982 for details). It is invariant under arbitrary smooth transformation. As it is associated with each point, we can analyze the distribution of pointwise dimension across points. Informally speaking, pointwise dimension of a point characterizes the how many dimension measure spans around a point. We have developed a statistical technique to estimate the pointwise dimension for a set of data points (Hidaka & Kashyap, 2013). Applying this, each point in the dataset is assigned with that a positive value of pointwise dimension.

Two-class classification

We performed classification analyses of demonstrator types based on each of those features described above. The performance of classification is treated as a measure of how characteristic each feature to discriminate demonstrator types. Specifically, for this two-class classification task, the imitator has exposed to a time series of a pair of angle and angular velocity, which reflects each movement demonstrated in the goal-achieved and the goal-failed condition. A part of each time series corresponding with the first 10 seconds was excluded from the training data, as of transient periods heavily depending on an initial state. The rest of the time series, corresponding with the last 50 seconds of the movement, was used as the training data for classification. We used one single long time series, since the system seems ergodic: namely, a time series with any starting initial states eventually converges to the same stationary near-periodic dynamical system.

In classification, each point in a given time series is treated as an independent sample, and the angle, angular velocity, mechanical energy, and pointwise dimension for each point was computed. The power spectrum of angle was computed with each part of time series within a moving time window of size 5 seconds. Then given a set of feature points as a training data, we used the Gaussian mixture model. This choice of the classifier is motivated by computational simplicity to construct two sample probability functions of a variable (feature). Denote these functions by $p_A(x)$ for the goal-achieved and $p_F(x)$ for the goal-failed demonstrator. Using these sample probability functions, the imitator asserts that a given new point x is of the goal-failed demonstrator if $p_F(x) > p_A(x)$, otherwise the imitator asserts that it is the point of the goal-achieved demonstrator. The classification accuracy is defined by the rate of correct response. For each feature, we reported the classification accuracy of the Gaussian mixture model with the minimum Bayesian information criterion.

Classification Results

Figure 3 shows the classification accuracy for each feature. As both the training and testing data contains the equally balanced number of samples for the two classes, the chance level for this classification was 50%. The classification accuracy with angle, angular velocity, and power spectrum was close to the change level. The accuracy with mechanical energy was approximately 95% significantly higher than the chance level. This result is as expected: the intention or motor control is a function of mechanical energy. We treat this accuracy on the basis of mechanical energy as an indicator of the best-possible accuracy in this classification task. Compared with this best-possible accuracy, the classification accuracy with pointwise dimension was approximately 90%, which was comparable to it. Note that pointwise dimension was computable with only a time series of angles of the pendulum observable by a naive imitator. This result suggests that pointwise dimension can be a potential characteristic to recognize intentions (controllers) behind observed movements only with observable data.

Simulation II: Action Completion Task

One of the key observation in the WT experiment is that the children can make an action to achieve the demonstrator's "goal" by just observing his incomplete action. As the children did not observe the complete action in their experiment, it needs to identify the observed incomplete action to the putative complete action at a certain level of abstraction. In order to explore this mechanism of the action completion task, we ask the question how the imitator

observing the goal-failed demonstration can reproduce an action substantially achieving the unobserved goal. As pointwise dimension was found reasonably characteristic of the intentions in Simulation I, we consider an extended use of pointwise dimension for the action completion task.

In the action completion task, exact identification of the intention is not necessarily required nor beneficial: as the imitator (e.g., child) does not necessarily have the same body as the demonstrator (e.g., adult), and the motor controllers for different bodies to meet the same goal may be different in general. In the action completion task, the imitator needs to identify two actions, which have similar goals but are different in body and latent motor control.

Action completion model

To this requirement, here we propose to use the similarity between two actions on the basis of their dynamic transition patterns in the DoF of the two action generating systems. In this model, the imitator observes an action and extracts dynamics of the DoFs in it as already shown in the recognition task of Simulation I. Next the imitator simulates a movement by its own body (a given specific pendulum) for each of a set of candidate controllers. Then the imitator makes an action which has the DoF dynamics the most similar with the extracted DoF dynamics of the demonstrated movement. In this way, this action completion model uses similarity in DoF dynamics, rather than similarity in apparent movement such as angle, angular velocity patterns.

Specifically, we suppose that the imitator is exposed to one time series of angles generated in the goal-failed condition (Figure 1C), which is sub-optimal for the swing-up-no-hit task. We assume that the imitator makes an action by choosing a controller (Equation 3) with goal angle G as the parameter. The other physical parameters, mass m and length l of the pendulum, are fixed $(m, l) = (1, 1)$ for both demonstrator and imitator in the same-pendulum condition, and are fixed $(m, l) = (4, 1)$ or $(m, l) = (1, 2)$ for the imitator in the two different-pendulum conditions. These conditions are designed to estimate the robustness of this action completion model to the physical difference of pendulums of the imitator and the demonstrator. Given the goal-failed action, the action completion task of the imitator is to choose likely controller (i.e., goal angle G) that generates a movement similar with the demonstrated movement in sense of their DoF dynamics. Each controller with the goal angles $0, 0.05, 0.1, \ldots, 0.9$ was analyzed as a candidate for action completion.

Similarity in DoF dynamics

In this study, the degree of freedom (DoF) dynamics of a system is defined by its temporal change in pointwise dimension estimated on the time series generated by the system. Specifically, a pointwise dimension estimator was constructed for a given demonstrated movement by the method proposed by Hidaka & Kashyap (2013), and this estimator was used to estimate a series of pointwise dimension for each of the demonstrated and candidate

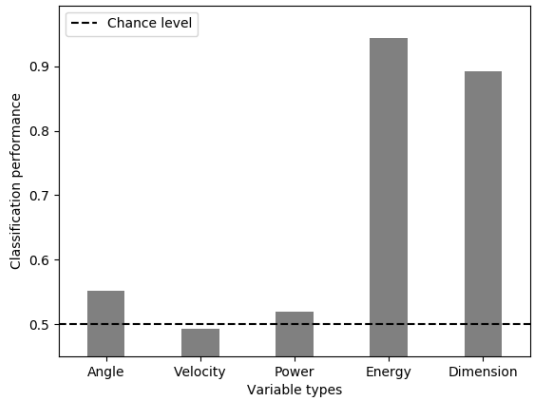

Figure 3: Results of classification tasks with several features

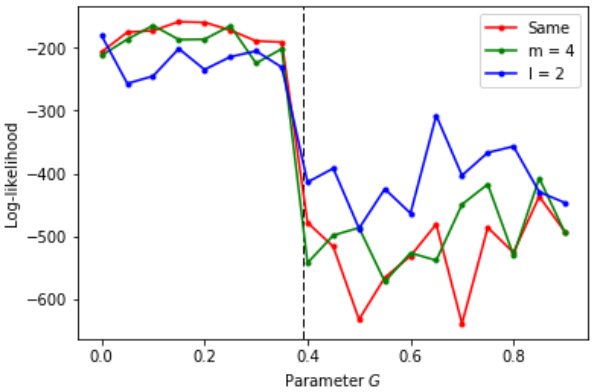

Figure 4: Results of the action completion task. The log-likelihood was computed by the action completion model. For the same-pendulum condition, the ground truth is $G = 0$. The vertical dashed line is the boundary of feasible angle $\theta_{\min} = \pi/8$. For the different-pendulum conditions, candidate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

movements. The constructed pointwise dimension estimator consists of multiple distributions, and each of distribution corresponds with a particular pointwise dimension. For each of the movements, we computed the matrix of transition probability from one distribution of a certain dimension to the other. The likelihood of a certain control parameter is defined by the multinomial distribution of a transition frequency generated by the controller with the given transition probability of pointwise dimension. This method is designed to abstract away difference between two systems in the absolute value of pointwise dimension at each step, and compute similarity in the temporal change in DoF.

Generation Results

For each of the set of goal angle G, the log-likelihood was computed based on similarity in the DoF dynamics between demonstrated and candidate movement (Figure 4). For each G, the figure shows the average log-likelihood over 20 sample actions with different initial values. A controller with some G with higher log-likelihood is more likely to be chosen as the produced action by the imitator. Figure 4 (red points) shows that the log-likelihood of goal angle G in the same-

pendulum condition, in case that both demonstrator and imitator control the same pendulum ($m = 1, l = 1$). As the goal angle in the goal-failed demonstrator was $G = 0$, it is the ground truth to be estimated. Although the log-likelihood did not take its highest at $G = 0$, it generally showed higher log-likelihood for one group $0 \leq G < \theta_{\min}$ than those for the other $G \geq \theta_{\min}$, where $\theta_{\min} \approx 0.39$ is the boundary angle between the feasible and infeasible one. These two groups of log-likelihoods on average were significantly different ($t(139) = 14.64, p < 0.01$). Thus, this result shows that, using similarity in the DoF dynamics, the imitator can generally differentiate the two latent types of candidate actions, which corresponding with the difference between the swing-up and swing-up-no-hit task. In other words, the imitator can reproduce some action which is similar with what the goal-failed demonstrator wanted to do but could not (i.e., swing it up through the infeasible region), if he/she were asked to perform in the swing-up task (no infeasible angle).

How dependent is this action completion model on the sameness of physical parameters of the demonstrator's and the imitator's pendulums? To see the dependence on the identical physical setting ($m = 1, l = 1$), we also analyzed the same action completion task where the imitator controls different pendulums with different physical parameters ($m = 4$ and $l = 1$, and, $m = 1$ and $l = 2$). In both cases, we successfully reproduced essentially similar results (the green and blue points in Figure 4) as that shown further same pendulum condition (the red points in Figure 4). The two groups of log-likelihoods on average were both significantly different ($t(139) = 14.10, t(139) = 8.78, p < 0.01$). That is, even with physically different pendulums, the imitator can differentiate the two general types of intentions (i.e., swing-up vs. swing-up-no-hit). Thus, by using the DoF dynamics as a basis of similarity between movements, the imitator can successfully abstract away the difference between their physical dissimilarity in the two pendulums. Note that, using two physically different pendulums, there is no more "ground truth" of the motor controller, producing a movement 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 of the demonstrator.

General Discussion

Inspired by the psychological experiment by Warneken & Tomasello (2006), we design our simulation as a minimal framework account for the mechanism for action recognition and action completion. We showed that the simulated imitator can discriminate the goal-failed and goal-achieved actions, whose movements apparently similar but the intentions and goals behind differ (Simulation I). Then we propose an action completion model that can make an action that generally similar with the action optimal for the swing-up task, by just observing the goal-failed action, which is sub-optimal to it (Simulation II).

Through these two simulations, we used the DoF dynamics in actions as a feature of the underlying motor controllers. Given this theoretical result, that the DoF dynamics is effective in both action recognition and completion, we predict that this feature would also play a crucial role in human action recognition and imitation, which will be tested in future work.

Here let us discuss the theoretical advantage of DoF dynamics (pointwise dimension) in the characterization of an action generating system over other approaches. The inverse kinematics approach (Wolpert et al., 2003), typically taken in robotics, needs the fully specified knowledge on the physical system including the motor controller in model parameter estimation. The Inverse reinforcement learning (IRL) approach (Ng & Russel, 2000) is a more general framework to estimate the unknown reward function by a given movement, under the assumption that the optimal controller trained by reinforcement learning generates the movement. It can well approximate the reward function for the highly rewarding states as such states are visited more frequently; otherwise, it may be poor to estimate the reward function for infrequent states. Therefore, IRL would not work well, if the most rewarding state is missing in the observed data – like the case of the goal-failed condition analyzed in this study.

In contrast to these previous approaches, the proposed model, at least in the minimal physical model such as a pendulum control task, can reasonably work for action completion task. We expect to extend the current work to a more complex action-generating system in future.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP16H05860, JP17H06713.

References

Astrom, K. J., & Furuta, K. (2000). Swinging up a pendulum by energy control. *Automatica*, *36*(2), 287-295.

Bernstein, N. A. (1996). *Dexterity and Its Development.* (M. L. Latash, & M. T. Turvey, Eds.) Psychology Press.

Cutler, C. D. (1993). A review of the theory and estimation of fractal dimension. In: *Nonlinear Time Series and Chaos*.

Doya, K. (1999). Reinforcement learning in continuous time and space. *Neural Computation*, *12*, 243-269.

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

Ng, A., & Russell, S. J. (2000). Algorithms for inverse reinforcement learning. *Proceedings of the Seventeenth International Conference on Machine Learning*, 663-670.

Tomasello, M. (2001). *The Cultural Origins of Human Cognition.* Harvard University Press.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, 1301-03.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Phil Trans B*, *358*(1431), 593-602.

Young, L.-S. (1982). Dimension, entropy, and Lyapunov exponents. *Ergodic Theory and Dynamical Systems*, *2*(1), 109-124.