# Context-dependent memory decay is evidence of effort minimization in motor learning: a computational study

## Ken Takiyama *

*Brain Science Institute, Tamagawa University, Tokyo, Japan*

Recent theoretical models suggest that motor learning includes at least two processes: error minimization and memory decay. While learning a novel movement, a motor memory of the movement is gradually formed to minimize the movement error between the desired and actual movements in each training trial, but the memory is slightly forgotten in each trial. The learning effects of error minimization trained with a certain movement are partially available in other non-trained movements, and this transfer of the learning effect can be reproduced by certain theoretical frameworks. Although most theoretical frameworks have assumed that a motor memory trained with a certain movement decays at the same speed during performing the trained movement as non-trained movements, a recent study reported that the motor memory decays faster during performing the trained movement than non-trained movements, i.e., the decay rate of motor memory is movement or context dependent. Although motor learning has been successfully modeled based on an optimization framework, e.g., movement error minimization, the type of optimization that can lead to context-dependent memory decay is unclear. Thus, context-dependent memory decay raises the question of what is optimized in motor learning. To reproduce context-dependent memory decay, I extend a motor primitive framework. Specifically, I introduce motor effort optimization into the framework because some previous studies have reported the existence of effort optimization in motor learning processes and no conventional motor primitive model has yet considered the optimization. Here, I analytically and numerically revealed that context-dependent decay is a result of motor effort optimization. My analyses suggest that context-dependent decay is not merely memory decay but is evidence of motor effort optimization in motor learning.

**Keywords: motor learning, neural network modeling, context-dependent memory decay, effort minimization, motor primitive**

## 1. INTRODUCTION

After a few years of not training in a previously learned sport, it is easy to forget how to move one's body in a manner that is suitable for the sport. This phenomenon is known as memory decay and is a well-known component of motor learning. Indeed, recent studies suggest that motor learning includes at least two processes: the minimization of movement error and motor memory decay (Scheidt et al., 2001; Smith et al., 2006; Hirashima and Nozaki, 2012).

The properties of error minimization have been extensively investigated using perturbation paradigms such as a force field paradigm (Shadmehr and Mussa-Ivaldi, 1994) or a visuomotor transformation paradigm (Krakauer et al., 1999). In these paradigms, subjects exhibit trial-by-trial adaptations to the novel environment, and the movement error between the desired and actual movements decreases in each trial. The learning effect of error minimization generalizes when kinematics (e.g., the movement direction) change (Thoroughman and Shadmehr, 2000; Donchin et al., 2003). For example, when trained in unimanual reaching movements toward a target direction $\theta$, the learning effects of the error minimization are partially available for

reaching movements toward other target directions. In other words, the availability of the learning effect depends on context (in this example, the context is equal to the target direction). In the following, I refer to the learning effects of error minimization as learning effects and to the context-dependent availability of learning effects as generalization.

Generalization can be modeled using a motor primitive framework (Thoroughman and Shadmehr, 2000; Donchin et al., 2003). In this framework, the activities of the motor primitive determine the motor commands, the recruitment pattern of the motor primitive is determined by a target direction, and the motor commands are updated to minimize movement error. The motor primitive framework can naturally reproduce the generalization because a group of motor primitives recruited in reaching movements toward a target direction $\theta$ overlaps with a group of primitives recruited in movements toward other directions, $\theta'(\neq \theta)$, which results in learning effects embedded in the motor primitives responsible for reaching movements toward $\theta$ are partially available in other reaching movements. Recent studies have revealed that the motor primitive framework can reproduce generalization in various situations, such as motor learning in a

force field (Thoroughman and Shadmehr, 2000; Donchin et al., 2003) or visuomotor rotation (Tanaka et al., 2009), motor learning of bimanual reaching movements (Yokoi et al., 2011), motor learning of unimanual reaching movements when the both target direction and shoulder posture change (Brayanov et al., 2012), the dependence of the generalization on error-feedback information (Taylor et al., 2012), and the generalization between reaching movements of the right hand and those of the left hand (Yokoi et al., 2014). Thus, the properties of error minimization or generalization in a wide range of situations can be explained by the motor primitive framework.

By contrast, the properties of memory decay have rarely been investigated. In the perturbation paradigms, the memory of the learning effects decays in each trial (Scheidt et al., 2000). In the following, I refer to the memory of learning effects as motor memory. Nearly all the models of motor primitives assume that the rate of motor memory decay is context independent (Thoroughman and Shadmehr, 2000; Scheidt et al., 2001; Donchin et al., 2003; Smith et al., 2006; Tanaka et al., 2009; Yokoi et al., 2011; Brayanov et al., 2012; Hirashima and Nozaki, 2012; Taylor et al., 2012; Takiyama and Okada, 2012a; Yokoi et al., 2014), i.e., the motor memory learned with reaching movements toward $\theta$ decays during the trials of the reaching movements at the same speed as during trials of reaching movements toward other directions, $\theta'(\neq \theta)$. In the motor primitive framework, the context-independent decay is equivalent to weight decay (Hirashima and Nozaki, 2012). The weight values of each motor primitive determine how the primitives contribute to generate motor commands, and the weight decay introduces a trial-by-trial decay of the magnitudes of the weight values (see the Weight Decay section in the Results for details). However, a recent behavioral experiment found that the decay rate of motor memory can be context dependent (Ingram et al., 2013). When trained in reaching movements toward a target direction $\theta$, the motor memory decayed faster during the trials of reaching movements toward $\theta$ than during trials of reaching movements toward other target directions $\theta'(\neq \theta)$. The authors of this study reported that a conventional model of motor learning could be fit well to their data with the assumptions of both generalization function and context-dependent memory decay. Notably, the authors heuristically introduced context-dependent decay to fit the conventional model to the actual data. Thus, the reason that memory decay is context dependent remains unclear. Although motor learning has conventionally been modeled as an optimization framework (Thoroughman and Shadmehr, 2000; Scheidt et al., 2001; Donchin et al., 2003; Smith et al., 2006; Tanaka et al., 2009; Yokoi et al., 2011; Brayanov et al., 2012; Hirashima and Nozaki, 2012; Taylor et al., 2012; Takiyama and Okada, 2012a,b; Yokoi et al., 2014), e.g., movement error minimization, no conventional optimization framework can reproduce the context-dependent memory decay, i.e., context-dependent memory decay raises the question of what is optimized in motor learning.

In the current study, I extend a motor primitive framework to reproduce the context-dependent decay of motor memory. Specifically, I introduce effort minimization, i.e., the minimization of motor command magnitude (Harris and Wolpert, 1998;
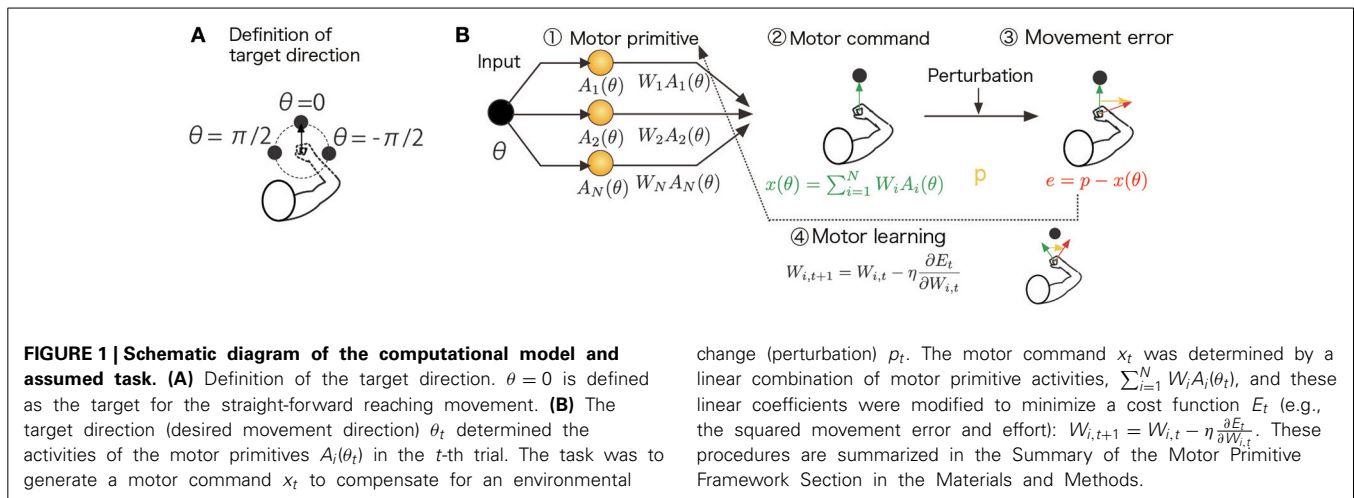
Fagg et al., 2002), into the framework because previous studies have reported the minimization during motor learning processes, such as adapting to a force field (Emken et al., 2007; Huang et al., 2012) and a visuomotor transformation (Huang and Ahmed, 2014) and because no conventional model of the motor primitive has yet considered effort minimization. Effort minimization is a widely accepted framework in motor control because the standard deviation of muscle activity is proportional to the muscle activity itself, which allows effort minimization to effectively minimize motor noise (Jones et al., 2002) or endpoint variance (Harris and Wolpert, 1998). Thus, it is natural to introduce effort minimization into conventional frameworks of motor primitives. Here, I analytically and numerically demonstrate that a motor primitive framework can reproduce context-dependent decay with the assumptions of both error and effort minimization.

## 2. MATERIALS AND METHODS
### 2.1. GENERAL FRAMEWORK

The present study focused on unimanual reaching movements in the horizontal plane toward radially distributed targets whose directions were defined as $\theta \in [-\pi, \pi]$ (**Figure 1A**). During each reaching movement, an unpredictable perturbation $p$, such as a force field (Shadmehr and Mussa-Ivaldi, 1994) or a visuomotor transformation (Krakauer et al., 1999), was imposed and yielded a movement error $e$ (**Figure 1B**). The aim of the task was to accurately reach to a given target by generating an additional motor command $x$ to compensate for the movement error, i.e., $e = p - x$. Here, following several previous studies, I assumed that $x$ represents the lateral force in the force adaptation (Yokoi et al., 2011; Brayanov et al., 2012; Yokoi et al., 2014) or the movement angle in the visuomotor rotation (Tanaka et al., 2009; Taylor et al., 2012) at the time of the peak velocity because these values are considered to represent the degree of adaptation in a feedforward controller, which many motor learning studies have focused on. That is, $p$ represents an applied lateral force in a curl force field paradigm (the units of $p$, $x$, and $e$ are newtons) and a rotated movement angle in a visuomotor rotation paradigm (the units of $p$, $x$, and $e$ are degrees). Notably, the movement error does not include the target direction. The motor primitive framework assumes that the motor commands of the baseline trials (before experiencing the adaptation paradigm) $x_b$ can be appropriately generated, i.e., $x_b = \tau_\theta$ in force field adaptation, where $\tau_\theta$ is the desired torque to achieve the reaching movements toward $\theta$, and $x_b = \theta$ in the visuomotor rotation adaptation. When the perturbation $p$ is applied and the compensatory motor command $x$ is generated, the baseline motor command is distorted as $x_b - p + x$, resulting in movement errors that can be written as $e = \tau_\theta - (x_b - p + x) = p - x$ in the force field adaptation and $e = \theta - (x_b - p + x) = p - x$ in the visuomotor rotation adaptation.

Following previous motor primitive models (Thoroughman and Shadmehr, 2000; Donchin et al., 2003; Tanaka et al., 2009; Yokoi et al., 2011; Brayanov et al., 2012; Taylor et al., 2012; Yokoi et al., 2014), we assumed that the recruitment pattern of the motor primitives or the motor primitive activities are determined by a target direction $\theta$ and the Gaussian

**FIGURE 1 | Schematic diagram of the computational model and assumed task. (A)** Definition of the target direction. $\theta = 0$ is defined as the target for the straight-forward reaching movement. **(B)** The target direction (desired movement direction) $\theta_t$ determined the activities of the motor primitives $A_i(\theta_t)$ in the $t$-th trial. The task was to generate a motor command $x_t$ to compensate for an environmental change (perturbation) $p_t$. The motor command $x_t$ was determined by a linear combination of motor primitive activities, $\sum_{i=1}^{N} W_i A_i(\theta_t)$, and these linear coefficients were modified to minimize a cost function $E_t$ (e.g., the squared movement error and effort): $W_{i,t+1} = W_{i,t} - \eta \frac{\partial E_t}{\partial W_{i,t}}$. These procedures are summarized in the Summary of the Motor Primitive Framework Section in the Materials and Methods.

$$A_i(\theta) = \exp\left(-\frac{1}{2\sigma_i^2} \|\theta - \varphi_i\|^2\right), \quad (1)$$

where the scaling parameter $\sigma_i = \sigma$ is independent of $i$; $\varphi_i$ is the preferred direction (PD) of the $i$-th primitive, which is randomly sampled from a uniform distribution in the range $[-\pi, \pi]$; $\|\theta\|$ is a periodic function $\|\theta\| = \|\theta + 2\pi\|$ such that $\|\theta\| = \theta$ for $-\pi \leq \theta < \pi$ ($\theta = 0$ is defined as the target for a straight-forward reaching movement, **Figure 1A**), and $i = 1, ..., N$, and $N$ is the number of primitives. The scaling parameter $\sigma$ controls the number of primitives responsible for a reaching movement toward $\theta$ because the $i$-th primitive is activated independent of $\theta$ when $\sigma$ is sufficiently large (i.e., the number of recruited primitives is large) and because the primitive shows its activity only in the case $\theta = \varphi_i$ when $\sigma$ is sufficiently small (i.e., the number of recruited primitives is small). The PD determines the reaching direction in which the $i$-th primitive shows its maximal activity. Because the PD was randomly sampled in each simulation run, the simulation results were variable across runs. The standard deviation of the simulation results is plotted in **Figures 3D–I**.

The compensatory motor command for the reaching movement toward $\theta$, $x(\theta)$, is a linear summation of the motor primitive activities $A_1(\theta), ..., A_N(\theta)$:

$$x(\theta) = \sum_{i=1}^{N} W_i A_i(\theta), \quad (2)$$

where the adaptable weight value $W_i$ determines how the $i$-th primitive contributes to the generation of the motor command. Each weight $W_i$ is modified by

$$W_{i,t+1} = W_{i,t} - \eta \frac{\partial E_t}{\partial W_{i,t}} \quad (3)$$

(gradient descent rule) at the $t$-th trial to reduce the cost function $E_t$, where the positive constant $\eta > 0$ denotes the learning rate. Conventional motor primitive frameworks assume that $E_t$ consists of the squared movement error $e_t^2$ and the sum of squared weight values $\sum_{i=1}^{N} W_{i,t}^2$, which leads to the learning rule of Equation (12) in the Results section or to context-independent decay of motor memory. Here, I suppose that the cost function consists of the squared movement error and the squared motor command amplitude (effort) $x_t^2$, which leads to the learning rule of Equation (15) in the Results section or to context-dependent decay of motor memory. When $E_t = \frac{1}{2}e_t^2 + \frac{\lambda_1}{2}\sum_{i=1}^{N} W_{i,t}^2 + \frac{\lambda_2}{2}x_t^2$, Equation (3) can be written as

$$W_{i,t+1} = (1 - \eta\lambda_1)W_{i,t} - \eta\lambda_2 x_t(\theta_t)A_i(\theta_t) + \eta e_t A_i(\theta_t), \quad (4)$$

where $\lambda_1$ and $\lambda_2$ are regularization parameters to determine the tradeoff between the minimization of the squared movement error, the minimization of the sum of the squared weight values, and the minimization of effort.

## 2.2. SUMMARY OF THE MOTOR PRIMITIVE FRAMEWORK

The motor primitive framework can be summarized as follows. Setting the parameters $\sigma$, $\eta$, $\lambda_1$, $\lambda_2$, $p$ to certain values, $e_0 = 0$, $W_0 = 0$, $\varphi_i$ to a randomly sampled value in the range $[-\pi, \pi]$, and the target direction $\theta_t \in [-\pi, \pi]$ to a certain value at the $t$-th trial yields the following:

(Determining recruitment pattern of motor primitives)

$$A_i(\theta_t) = \exp\left(-\frac{1}{2\sigma^2}\|\theta_t - \varphi_i\|^2\right). \quad (5)$$

(Generation of a motor command) $x_t(\theta_t) = \sum_{i=1}^{N} W_{i,t} A_i(\theta_t).$ (6)

(Observation of a movement error) $e_t = p_t - x_t(\theta_t).$ (7)

(Update of linear coefficients)

$$W_{i,t+1} = (1 - \eta\lambda_1)W_{i,t} - \eta\lambda_2 x_t(\theta_t)A_i(\theta_t) + \eta e_t A_i(\theta_t). \quad (8)$$

Notably, with the assumptions of the motor primitive frameworks and the gradient descent as the learning rule, these procedures are exact and without approximation. This motor primitive framework can explain trial-dependent changes in the movement

error and generalization (Thoroughman and Shadmehr, 2000; Donchin et al., 2003).

## 2.3. GENERALIZATION FUNCTION

Learning effects trained with reaching movements toward $\theta_t$ were generalized to untrained reaching movements toward $\theta$ through a generalization function $G(\theta_t, \theta)$, where $\theta_t$ and $\theta$ were assumed to be the target directions in the training and test trials, respectively. Test trials indicate the trials to probe learning effects during or after the training trials. Equation (4) indicates the learning process during the training of reaching movements toward $\theta_t$. The degree of generalization from the reaching movements toward $\theta_t$ to other reaching movements toward $\theta$ can be derived by multiplying $A_i(\theta)$ and summing across all values of $i$ in Equation (4):

$$x_{t+1}(\theta) = x_t(\theta) + \eta \mathcal{G}(\theta_t, \theta)e_t, \qquad (9)$$

where the generalization function is defined as $\mathcal{G}(\theta_t, \theta) = \sum_{i=1}^{N} A_i(\theta_t)A_i(\theta)$ and both $\lambda_1$ and $\lambda_2$ are set to 0 for simplicity (cases in which $\lambda_1 \neq 0$ or $\lambda_2 \neq 0$ are discussed in the Results section). The generalization function $G(\theta_t, \theta)$ is defined as the inner product of the recruitment pattern of motor primitives during reaching toward $\theta_t$ and those for reaching toward $\theta$. Larger overlaps between the two recruitment patterns result in greater generalization. In addition, the generalization function is dependent only on $\sigma$ (see Equation (10) described below) and is independent of other parameters, including the perturbation $p$.

More intuitively, when the number of primitives $N$ is sufficiently large and $\eta$ is divided by $N$ (most conventional motor primitive models assume these conditions), the generalization function can be approximated as

$$\frac{\eta}{N}\mathcal{G}(\theta_t, \theta) \simeq \eta\sqrt{\pi\sigma^2} \exp\left(-\frac{1}{4\sigma^2}||\theta - \theta_t||^2\right), \qquad (10)$$

which indicates that the closer the target direction in the test trials ($\theta$) is to the direction in the training trials ($\theta_t$), the greater the generalization. In addition, the generalization function depends only on $\sigma$. Although this derivation of Equation (10) requires a sufficiently large $N$, $N$ does not significantly influence the shape of $G(\theta_t, \theta)$ (Takiyama and Okada, 2012a). Thus, Equation (10) permits an intuitive understanding of the results of the current study. Furthermore, this type of generalization function suitably explains the actual data from a wide range of motor learning experiments (Thoroughman and Shadmehr, 2000; Donchin et al., 2003; Tanaka et al., 2009; Yokoi et al., 2011; Brayanov et al., 2012; Taylor et al., 2012; Yokoi et al., 2014).

## 2.4. STATE SPACE MODELING

Although the motor primitive framework assumed trial-by-trial variations in the weight value, $W$, it was possible to model only the trial-by-trial variation of the motor command $x$. In conventional state space models, motor learning is modeled as an optimization process regarding motor commands:

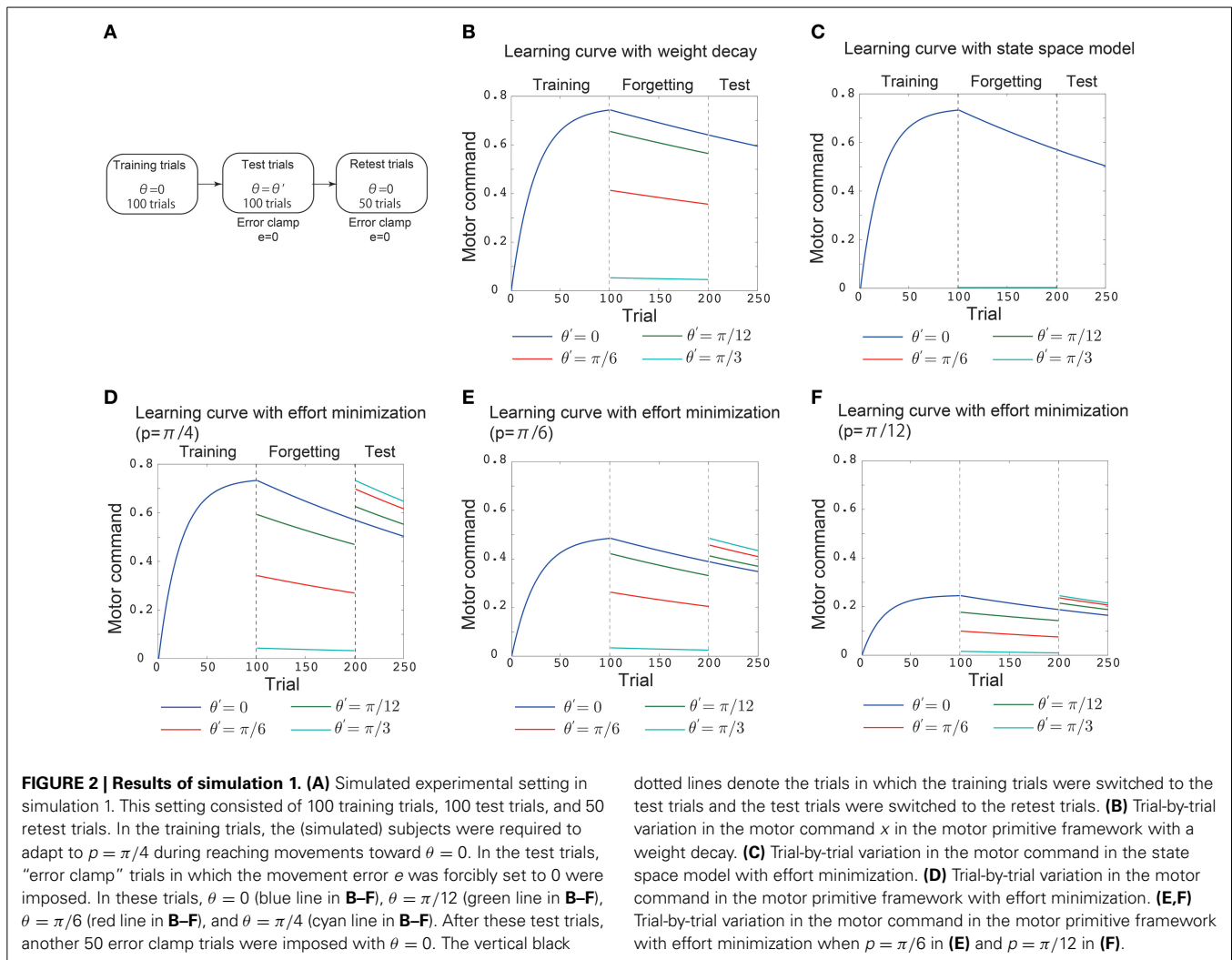$$x_{t+1} = x_t - \eta\frac{\partial E_t}{\partial x_t}, \qquad (11)$$

where $E_t$ is a cost function at the $t$-th trial, e.g., the squared movement error. Notably, Equation (9) is equivalent to Equation (11) when $\theta = \theta_t$ and $\sqrt{\pi\sigma^2}\eta$ is redefined as $\eta$. Thus, this state space model can be viewed as a motor learning model that focuses only on the trial-by-trial variation of the motor commands without assuming any generalization function. A previous study demonstrated that the state space model fits the data well if context-dependent decay and the generalization function are heuristically introduced into this model (Ingram et al., 2013), i.e., $x_{t+1}(\theta) = (1 - \eta\lambda_2\mathcal{G}_1(\theta_t, \theta))x_t(\theta) + \eta\mathcal{G}_2(\theta_t, \theta)e_t$ ($\mathcal{G}_1$ and $\mathcal{G}_2$ are some functions).

## 2.5. NUMERICAL SIMULATION

The learning rates in the motor primitive framework and state space modeling, i.e., $\eta$ in Equations (12) and (15) and $\eta$ in Equation (14), were set to 0.5 and 0.04, respectively. The rates of forgetting in weight decay ($\lambda_1$ in Equation 12) and in state space modeling ($\lambda_2$ in Equation 14) were set to 0.0015/$\eta$. The rate of forgetting in effort minimization in a motor primitive framework ($\lambda_2$ in Equation 15) was set to 0.03/$\eta$. The number of motor primitives $N$ was set to 100, and the perturbation $p$ was set to $\pi/4$. The three types of models considered in the current study, i.e., the motor primitive framework with weight decay, the state space model with effort minimization, and the motor primitive with effort minimization, produced nearly identical learning curves using these parameters (**Figure 3D**). To investigate the parameter dependencies, $p$ was set to $\pi/6$ or $\pi/12$ in **Figures 2E,F**, respectively, and $\eta$ and $\lambda_2$ were set to 0.8 and 0.01 in **Figures 3H,I**.

## 3. RESULTS

I analytically and numerically investigated a condition in which the decay of motor memory and the availability of learning effects were context dependent. The present study assumed a task involving unimanual reaching movements toward radially distributed target directions in the horizontal plane (**Figure 1A**; detailed descriptions of the following framework, analytical calculations, and each parameter value were provided in the Materials and Methods section). The goal of the task was to reach to the target accurately in a situation in which executed movements were perturbed by changes in the environment $p$, e.g., the external force generated by a manipulandum (Shadmehr and Mussa-Ivaldi, 1994) or visuomotor transformation (Krakauer et al., 1999). The motor command $x$ to compensate for a perturbation was modeled by the summation of the activities of the motor primitives as $x(\theta) = \sum_{i=1}^{N} W_i A_i(\theta)$, where the adaptable $W_i$ represents how the $i$-th primitive contributes to the production of the motor command and $A_i(\theta)$ is the activity of the $i$-th primitive determined by the target direction. The current study assumed that the recruitment pattern of motor primitives was determined by the target direction. The movement error in the $t$-th trial can thus be expressed as $e_t = p_t - x_t$. To minimize the squared movement error, the squared sum of the weight values, and the squared motor

**FIGURE 2 | Results of simulation 1. (A)** Simulated experimental setting in simulation 1. This setting consisted of 100 training trials, 100 test trials, and 50 retest trials. In the training trials, the (simulated) subjects were required to adapt to $p = \pi/4$ during reaching movements toward $\theta = 0$. In the test trials, "error clamp" trials in which the movement error $e$ was forcibly set to 0 were imposed. In these trials, $\theta = 0$ (blue line in **B–F**), $\theta = \pi/12$ (green line in **B–F**), $\theta = \pi/6$ (red line in **B–F**), and $\theta = \pi/4$ (cyan line in **B–F**). After these test trials, another 50 error clamp trials were imposed with $\theta = 0$. The vertical black dotted lines denote the trials in which the training trials were switched to the test trials and the test trials were switched to the retest trials. **(B)** Trial-by-trial variation in the motor command $x$ in the motor primitive framework with a weight decay. **(C)** Trial-by-trial variation in the motor command in the state space model with effort minimization. **(D)** Trial-by-trial variation in the motor command in the motor primitive framework with effort minimization. **(E,F)** Trial-by-trial variation in the motor command in the motor primitive framework with effort minimization when $p = \pi/6$ in **(E)** and $p = \pi/12$ in **(F)**.

command amplitude (effort), $W_i$ was modified in each trial (Equation 4).

### 3.1. WEIGHT DECAY

Conventional models of motor primitives assume weight decay (Thoroughman and Shadmehr, 2000; Donchin et al., 2003; Tanaka et al., 2009; Yokoi et al., 2011; Brayanov et al., 2012; Hirashima and Nozaki, 2012; Takiyama and Okada, 2012a; Taylor et al., 2012; Yokoi et al., 2014), i.e., the cost function $E_t$ to be minimized is defined as a weighted sum of the squared error and the squared sum of weight values $E_t = \frac{1}{2}e_t^2 + \frac{\lambda_1}{2}\sum_{i=1}^{N} W_{i,t}^2$, where a regularization parameter $\lambda_1$ determines the trade-off between error minimization and weight value minimization. The learning rule can thus be written as

$$W_{i,t+1} = (1 - \eta\lambda_1)W_{i,t} + \eta e_t A_i(\theta_t), \qquad (12)$$

where $\lambda_1$ indicates the forgetting rate and $\eta$ denotes the learning rate: The larger $\lambda_1$ is, the faster the motor memory decays, and the larger $\eta$ is, the faster the movement error is minimized. In addition, the larger the $i$-th primitive is activated, the faster the

motor memory can be embedded in the $i$-th primitive because the learning rate $\eta$ is modulated by $A_i$ in Equation (12). However, the forgetting rate is not modulated by $A_i$, which results in motor memory decays in the trial when the $i$-th primitive is not activated. The assumed situation in this equation was an adaptation process in which the subjects adapted to a perturbation with reaching movements toward $\theta_t$. Equation (12) can be rewritten as a recursive equation of $x$:

$$x_{t+1}(\theta) = (1 - \eta\lambda_1)x_t(\theta) + \eta\mathcal{G}(\theta_t, \theta)e_t, \qquad (13)$$

where $\mathcal{G}(\theta_t, \theta) = \sum_{i=1}^{N} A_i(\theta_t)A_i(\theta)$ is the generalization function (detailed derivation and description are provided in the Generalization Function section in the Materials and Methods). This equation indicates how the learning effects trained during reaching toward $\theta_t$ are generalized to the untrained reaching movements toward $\theta$ in each trial. The generalization function indicates that the learning effects can be largely observed in test trials when tested movement direction $\theta$ is close to the trained movement direction $\theta_t$. Although this generalization function can be fit well to a wide range of experimental data (Thoroughman
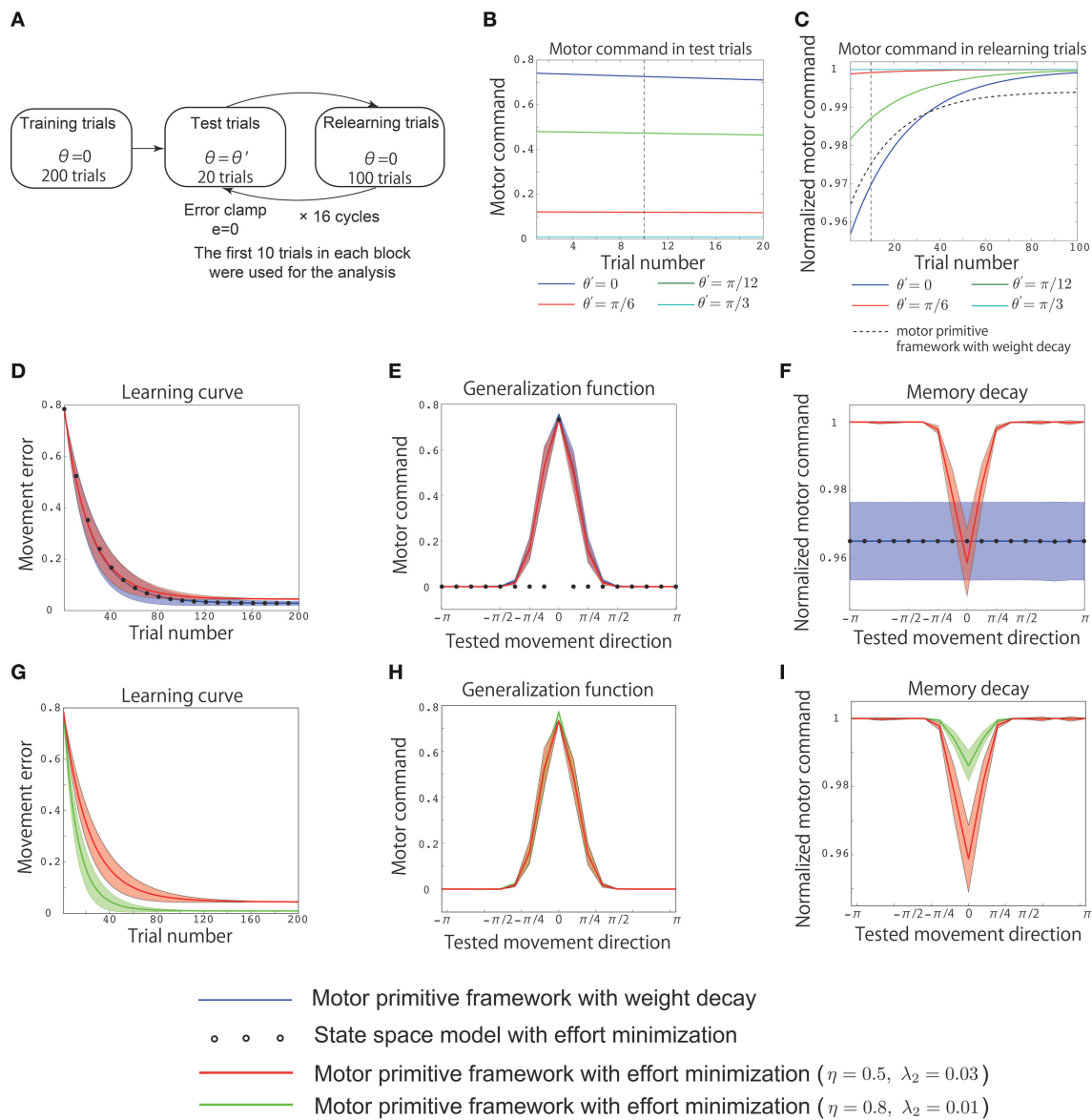
**FIGURE 3 | Results of simulation 2. (A)** Simulated experimental setting for simulation 2. Training trials were conducted for 200 trials with $\theta = 0$. After the training, 20 test trials with $\theta' = -\pi + 2\pi\frac{k}{K}$ and 100 relearning trials with $\theta = 0$ were alternately simulated for $K$ cycles, where $K = 16$, and the integer $k$ was pseudorandomly sampled from the range $[0, K - 1]$ to assume different values in each cycle. In the test trials, the movement error was forcibly set to 0 assuming error clamp trials. **(B)** Trial-by-trial variation of the motor command $x$ in the test trials with the motor primitive framework with effort minimization, $\theta' = 0$ (blue line), $\theta' = \pi/12$ (green line), $\theta' = \pi/6$ (red line), and $\theta' = \pi/3$ (cyan line). The vertical black dotted line indicates the 10th test trial. **(C)** Trial-by-trial variation of the motor command in the relearning trials with the motor primitive framework with effort minimization. The black dotted line denotes the trial-by-trial variation of the motor command in the relearning trials with the motor primitive framework with weight decay (independent of $\theta'$). The vertical black dotted line indicates the 10th relearning trial. **(D)** Learning curves. The horizontal axis indicates the trial number, and the vertical axis denotes the movement error $e_t$. The red line, blue line, and circle denote the trial-by-trial variation of $e_t$ averaged across 20 simulation runs in the motor primitive framework with effort minimization, the framework with weight decay, and the state space model with effort minimization, respectively. The

red and blue shaded areas indicate the standard deviations of the learning curves in the 20 simulation runs in the motor primitive framework with effort minimization and those with weight decay, respectively. After 200 trials, the motor commands $x_{200}(0)$ converged to $x_0$ in all three models. **(E)** Generalization function averaged across 20 simulation runs. The horizontal axis indicates the tested movement direction $\theta'$, and the vertical axis indicates the motor command $x(\theta')$. To draw the generalization function, I averaged the motor commands of the initial 10 test trials in each cycle (the 10th test trial is indicated by the vertical black dotted line in **B**). The red and blue shaded areas indicate the standard deviation of the generalization function in the 20 simulation runs in the motor primitive frameworks with effort minimization and weight decay, respectively. **(F)** Context-dependent decay investigated in the relearning phase with $\theta' = 0$. The horizontal axis indicates the tested movement direction $\theta'$, and the vertical axis indicates the normalized motor command in relearning trials. The normalized motor command was defined as $\frac{1}{10}\sum_{t=1}^{10} x_{\text{relearning},t}(0)/x_0$, where $x_{\text{relearning},t}(0)$ is a motor command at the $t$-th relearning trial in each cycle, i.e., I averaged the motor commands of the initial 10 relearning trials in each cycle (the 10th relearning trial is indicated by the vertical black dotted line in **C**) and divided the averaged value by $x_0$. The

*(Continued)*

and Shadmehr, 2000; Donchin et al., 2003; Tanaka et al., 2009; Yokoi et al., 2011; Brayanov et al., 2012; Taylor et al., 2012; Yokoi et al., 2014), the decay of motor memory, the first term in the right side of Equation (13), is independent of the target directions $\theta$ and $\theta_t$ (i.e., the context-independent memory decay), in contrast to the results of a previous behavioral experiment (Ingram et al., 2013). Thus, the motor primitive framework with weight decay does not reproduce context-dependent memory decay.

## 3.2. EFFORT MINIMIZATION IN A STATE-SPACE MODEL

In addition to the motor primitive framework, the state space model is another widely used model of motor learning (Scheidt et al., 2001; Smith et al., 2006; Lee and Schweighofer, 2009). The state space model focuses on the trial-by-trial variation of the motor command $x$ rather than the trial-by-trial variation of the weight value $W$. Thus, weight decay cannot be assumed in the state space model because this model does not assume any weight value. Memory decay was modeled as effort minimization in the state space model using the cost function $E_t = \frac{1}{2}e_t^2 + \frac{\lambda_2}{2}x_t^2$, i.e., a weighted sum of squared error and effort. Optimizing this cost function led to the minimization of movement error and effort. The learning rule can be written as

$$x_{t+1}(\theta_t) = (1 - \eta\lambda_2)x_t(\theta_t) + \eta e_t, \qquad (14)$$

where the memory decay and the availability of learning effects are context independent, also in contrast to the results of previous behavioral experiments (Thoroughman and Shadmehr, 2000; Donchin et al., 2003; Ingram et al., 2013). Although the state space model can be fit well to actual data by heuristically introducing context-dependent memory decay and a generalization function (Ingram et al., 2011, 2013), such modeling leaves the question of what is optimized in motor learning. While more complicated versions of state space models can be considered (Smith et al., 2006; Lee and Schweighofer, 2009; Takiyama et al., 2009; Takiyama and Okada, 2011), context-dependent memory decay cannot be reproduced. Thus, the state space model is a powerful model in some cases (Scheidt et al., 2001; Smith et al., 2006; Lee and Schweighofer, 2009), but is not suitable for modeling and interpreting the context-dependent decay of motor memory and generalization in an optimization framework.

## 3.3. EFFORT MINIMIZATION IN THE MOTOR PRIMITIVE FRAMEWORK

In the previous two models, the parameter being optimized differed: the weight value was minimized in the conventional motor primitive framework, while effort was minimized in the state space model. Although the functional roles of weight decay in motor learning remain unclear (cf. Hirashima and Nozaki, 2012), effort minimization has been reported to be effective in minimizing motor noise (Jones et al., 2002) and endpoint variance

(Harris and Wolpert, 1998). Furthermore, some previous studies have reported that not only movement error but also effort is minimized in the adaptation to a force field (Emken et al., 2007; Huang et al., 2012) and a visuomotor transformation (Huang and Ahmed, 2014). Thus, I propose a motor primitive framework with effort minimization. When $E_t$ is defined as $E_t = \frac{1}{2}e_t^2 + \frac{\lambda_2}{2}x_t^2$, motor learning can be modeled as

$$W_{i,t+1} = W_{i,t} - \eta\lambda_2 x(\theta_t)A_i(\theta_t) + \eta e_t A_i(\theta_t). \qquad (15)$$

The second term in the right side of Equation (15) is a memory decay that can be modulated by $x(\theta_t)$ and $A_i(\theta_t)$; the larger the $i$-th primitive is activated, the faster the motor memory embedded in the primitive decays. Equation (15) yields a recursive equation for motor commands:

$$x_{t+1}(\theta) = (1 - \eta\lambda_2\mathcal{G}(\theta_t, \theta))x_t(\theta) + \eta\mathcal{G}(\theta_t, \theta)e_t, \qquad (16)$$

where both memory decay and error minimization are context dependent. When trained with reaching movement toward $\theta_t$ and tested with movements toward $\theta$, the motor memory decays faster in the test trials when the tested movement direction is close to the trained movement direction because the forgetting rate $\eta\lambda_2\mathcal{G}(\theta_t, \theta)$ is maximal when $\theta = \theta_t$. This result is consistent with results of a previous experiment (Ingram et al., 2013). Taken together, effort minimization in a motor primitive framework is the only candidate that can simultaneously reproduce context-dependent memory decay and generalization.

## 3.4. SIMULATION 1: VALIDATION OF THE ANALYTICAL CONSIDERATION

To validate the analytical considerations, I conducted numerical simulations (**Figures 2**, **3**). The first simulation consisted of 100 training trials, 100 test trials, and 50 retest trials (**Figure 2A**). In the training trials, the (simulated) subjects were required to adapt to $p = \pi/4$ in reaching movements toward $\theta = 0$. The test trials were "error clamp" trials in which the movement error $e$ was forcibly set to 0 to enable a clear discussion of learning effects by excluding any adaptation process (Scheidt et al., 2000). These test trials were simulated to investigate how the learning effects trained in training trials are generalized to reaching movements toward $\theta = 0$ (blue line), $\theta = \pi/12$ (green line), $\theta = \pi/6$ (red line), or $\theta = \pi/4$ (cyan line). Notably, the error clamp trials also enabled a clear discussion of forgetting processes because no adaptation can be assumed to occur during the trials. Thus, the memory of learning effects trained in the training trials decayed during the test trials depending on $\theta$. To investigate how the target direction (context) affected the decay of the motor memory, another 50 error clamp trials were imposed with $\theta = 0$ after the test trials (**Figure 2A**).

The above simulated experimental setting enabled an investigation of whether the decay of motor memory and the availability of learning effects are context dependent or context independent. If the availability of learning effects is context dependent, the learning effects trained with $\theta = 0$ in the training trials will be generalized to other reaching movements in the test trials, and the motor command toward the other target directions $x(\pi/12)$, $x(\pi/6)$, and $x(\pi/4)$ will not be 0 in those trials. By contrast, if the availability of learning effects is context independent, then no generalization should be observed, and $x(\pi/12)$, $x(\pi/6)$, and $x(\pi/4)$ are 0 in the test trials. If the memory decay is context independent, the motor commands in the retest trials should be independent of the target direction in the test trials, i.e., the blue, green, red, and cyan lines that indicate $x(0)$ should converge to the same value in the retest trials. This result would indicate that the motor memory decays at the same rate in the test trials independent of the target direction. However, if the memory decay is context dependent, then the motor commands in the retest trials should depend on the target direction in the test trials, and the blue, green, red, and cyan lines that indicate $x(0)$ should diverge to different values in the retest trials. This result would indicate that the motor memory decays with different rates in the test trials depending on the target direction.

In the motor primitive framework with weight decay Equation (12), the availability of learning effects is context dependent (**Figure 2B**, the blue, green, red, cyan lines in the test trials). However, the memory decay is context independent (**Figure 2B**, all the lines converged to the same value in retest trials). In the state space model with effort minimization Equation (14), neither the availability of learning effects nor the memory decay is context dependent (**Figure 2C**). Thus, these conventional frameworks are not sufficient to reproduce both generalization and context-dependent memory decay.

By contrast, the motor primitive framework with effort minimization Equation (15) reproduced both generalization and the context-dependent memory decay (**Figure 2D**). Furthermore, these generalization and context-dependent decay were independent of the magnitude of perturbation (**Figures 2E,F**). Thus, my analytical considerations were validated; context-dependent memory decay is a result of effort minimization in the motor primitive framework.

### 3.5. SIMULATION 2: REPRODUCTION OF PREVIOUS EXPERIMENTAL DATA

Next, I reproduced the results of a previous behavioral experiment (Ingram et al., 2013) using the assumptions of a similar experimental setting (**Figure 3**). The training involved 200 trials with $\theta = 0$. After these trials, the motor command $x_{200}(0)$ converged to $x_0$ (**Figure 3D**). After the training, 20 test trials with $\theta' = -\pi + 2\pi \frac{k}{K}$ and 100 relearning trials with $\theta = 0$ were alternately simulated for $K$ cycles in which the integer $k$ was peudorandomly sampled from the range $[0, K-1]$ to take different value in each cycle and $K$ was 16. Once the integer $k$ was sampled, $k$, or $\theta'$ was fixed across the 20 test trials. The movement error was forcibly set to 0 in the test trials (error clamp trials). These test trials enabled to investigate the generalization. If the availability of the learning effects is context independent, motor commands

in test trials will be 0 except for the case when $\theta' = 0$. By contrast, the availability of the learning effects is context dependent, the learning effects trained in training or relearning trials will be generalized to other reaching movements in test trials. Because the magnitude of the motor commands (motor memory) decayed during the test trials (**Figure 3B**), I averaged the motor commands in the initial 10 test trials to obtain the generalization function shown in **Figures 3E,H** to eliminate the effects of decay on the function.

After the test trials, 100 relearning trials with $\theta = 0$ were imposed to investigate the context dependence of the memory decay. If the decay rate of the memory was independent of $\theta'$ in the test trials, the motor commands in the relearning trials should have the same values independent of $\theta'$. By contrast, if the decay rate of the memory depended on $\theta'$, the motor commands in the relearning trials should have different values depending on $\theta'$. Because the motor commands converged to the same value during the relearning trials (**Figure 3C**), I averaged the motor commands in the initial 10 relearning trials to obtain the context-dependent memory decay in **Figures 3F,I** to eliminate the effects of relearning on the function of the context-dependent memory decay. In addition, based on a previous study (Ingram et al., 2013), the motor commands in relearning trials were denoted after normalizing by $x_{200}(0)$ in the training trials (**Figures 3C,F,I**).

Similar to the previous section, Simulation 1: Validation of the Analytical Consideration, neither the motor primitive framework with weight decay nor the state space model with effort minimization reproduced the context-dependent memory decay in an experimental setting similar to that of the previous study (Ingram et al., 2013) (blue line and circles in **Figures 3E,F**). By contrast, the motor primitive framework with effort minimization reproduced both the context-dependent memory decay and the generalization in this experimental setting (red line in **Figures 3E,F**). Furthermore, these results were not sensitive to parameter changes (**Figures 3G–I**); although the magnitude of the memory decay was reduced when $\lambda_2$ was small, the context dependencies of the memory decay and the generalization were invariant. Thus, this framework was able to reproduce the generalization and context-dependent decay within an experimental setting similar to that of the previous study (Ingram et al., 2013). Taken together, the context-dependent decay and the generalization were the results of optimizing the effort and movement error in the motor primitive framework.

## 4. DISCUSSION

In the current study, I revealed that context-dependent decay is evidence of effort minimization in motor learning by extending a motor primitive framework. The conventional motor primitive framework succeeded in reproducing the availability of the learning effects trained with reaching movements toward $\theta$ in other reaching movements (Thoroughman and Shadmehr, 2000; Donchin et al., 2003; Tanaka et al., 2009; Yokoi et al., 2011; Brayanov et al., 2012; Taylor et al., 2012; Yokoi et al., 2014). Although the conventional models of motor learning support the existence of memory decay in motor learning processes (Thoroughman and Shadmehr, 2000; Scheidt et al., 2001; Donchin et al., 2003; Smith et al., 2006; Tanaka et al., 2009;

Yokoi et al., 2011; Brayanov et al., 2012; Hirashima and Nozaki, 2012; Takiyama and Okada, 2012a; Taylor et al., 2012; Yokoi et al., 2014), these conventional models assume that the rate at which memory decays is independent of the target direction or assume context-independent decay of motor memory. Context-independent decay is equivalent to the minimization of weight values (weight decay) in a motor primitive framework Equations (12) and (13) and effort minimization in a state space model Equation (14). Nevertheless, a recent behavioral experiment (Ingram et al., 2013) reported that the rate of memory decay is context dependent. Although motor learning has been modeled by an optimization framework, context-dependent decay raises the following question: what is optimized in motor learning? Here, based on the results of recent behavioral experiments (Emken et al., 2007; Huang et al., 2012; Huang and Ahmed, 2014), I introduced effort minimization into a motor primitive framework and concluded that context-dependent decay results from effort minimization in the motor primitive framework Equations (15) and (16), **Figures 2D**, **3E,F**). Furthermore, this context dependence of the motor memory decay was not sensitive to the parameters $\eta$, $\lambda_2$, and $p$ (**Figures 2E,F**, **3G–I**). This finding permits the seamless connection of conventional models of motor learning with the recent finding of context-dependent decay.

The ability of effort minimization to reproduce context-dependent decay is related to activity-dependent memory decay. In the error clamp trials, Equation (15) can be written as $W_{i,t+1} = W_{i,t} - \eta\lambda_2 x_t A_i(\theta_t)$, which indicates that the more the $i$-th primitive is activated in the $t$-th trial, the faster the memory embedded in the primitive decays. In other words, if the $i$-th primitive is not activated in a trial, the memory embedded in the primitive can be completely maintained during the error clamp trials. When trained with $\theta = 0$ and test trials imposed with $\theta = \pi/4$ (cyan line in **Figures 2D–F**), the recruitment patterns differed greatly between the training and test trials, and thus the motor memories embedded in the motor primitives that were activated when $\theta = 0$ were maintained in the test trials with $\theta = \pi/4$. Thus, after the test trials with $\theta = \pi/4$, the motor memory was maintained, and the magnitudes of the motor commands in the retest trials were not significantly different from the motor commands in the latter half of the training trials (**Figures 2D–F**). When trained with $\theta = 0$ and test trials imposed with $\theta = 0$ (blue line in **Figures 2D–F**), the same primitives were activated during the training and test trials, which led to faster forgetting of the motor memory in the test trials. Thus, after the test trials with $\theta = 0$, the motor memory decays and the magnitudes of the motor commands in the retest trials were smaller than the motor commands in the latter half of the training trials (**Figures 2D–F**). By contrast, weight decay entails activity-independent memory decay. In the error clamp trials, Equation (12) can be written as $W_{i,t+1} = (1 - \lambda_1)W_{i,t}$, which indicates that the motor memory embedded in the $i$-th primitive decays trial-by-trial independent of the activity of the $i$-th primitive or the target direction in the test trials (**Figure 2B**). Hence, activity-dependent memory decay is responsible for the association of effort minimization with the context-dependent decay of motor memory.

Functional roles of effort minimization in motor learning needs to be discussed. One possible functional role of effort minimization is to reduce mean squared movement error when a motor command includes signal-dependent noise (the noise whose standard deviation is proportional to its mean value), i.e., $y = x + \xi$, where $y$ is a motor command that includes motor noise, $x$ is a noiseless motor command, and $\xi$ is a motor noise with a mean of 0 and a variance of $\lambda_2 x^2$. In this case, the mean squared error can be written as follows: $\langle e_t^2 \rangle = (p_t - \langle y_t \rangle)^2 + \langle (y_t - \langle y_t \rangle)^2 \rangle = e_t^2 + \lambda_2 x_t^2$, where $\langle f(\xi_t) \rangle$ indicates $f(\xi_t)$ averaged across all possible realizations $\langle f(\xi_t) \rangle = \int d\xi_t p(\xi_t) f(\xi_t)$. This equation indicates that the mean squared movement error consists of constant error, $(p_t - x_t)^2$ and variable error $\langle (y_t - x_t)^2 \rangle = \lambda_2 x_t^2$ (Schmidt and Lee, 1988). The constant error represents the bias of the motor command. The smaller (larger) the motor command is down-regulated (up-regulated), e.g., $y_t = p_t - b + \xi_t$ ($y_t = p_t + b + \xi_t$), where $b$ is bias, the larger the constant error. By contrast, the variable error represents how variable the motor command is in each trial. The more variable the motor command, the larger the variable error. When the motor noise is signal dependent, effort minimization can decrease the trial-by-trial variability such that effort minimization decreases the variable error and the mean squared movement error.

One limitation of this study must be discussed. Although a previous behavioral experiment reported that both context-dependent decay and the generalization function can be modeled by Gaussian functions, the width of the function was wider for context-dependent decay than for the generalization function (Ingram et al., 2013). In Equation (16), the width of the Gaussian was the same for the two functions; the reason for the difference in the widths reported in the previous study remains unclear. A possible solution is the introduction of a multi-rate concept (Smith et al., 2006) into the motor primitive framework (Tanaka et al., 2012). In this concept, the motor memory consists of at least two processes, i.e., fast and slow processes in which, roughly speaking, the fast processes, including a large forgetting and a large learning rate, are involved when the movement error is large (e.g., the early part of the training trials) and slow processes, including a small forgetting and a small learning rate, are involved when the movement error is small (e.g., the latter part of the training trials). In the previous study (Ingram et al., 2013), generalization functions were investigated using error-clamp trials in which the movement error was forcibly set to 0. By contrast, for investigating context-dependent decay, error clamp trials were not used, which indicates that the movement error was not 0. If a larger number of primitives are activated in a fast process compared to a slow process, i.e., if $\sigma$ is larger in the motor primitives associated with the fast process than in those associated with the slow process, the difference in the widths of the context-dependent decay and the generalization function can be reproduced. Another possible solution is to determine the recruitment pattern of motor primitives based on the movement error (Takiyama et al., 2013). In this model, the recruitment patterns of the motor primitives are different between the trials used to investigate the context-dependent decay (i.e., when the movement error is large) and the trials used to investigate the generalization function (i.e., when the movement error is small) in the experimental setting, which results in a difference in $\sigma$ between the recruited primitives that enables the reproduction of the difference in the

widths of the context-dependent decay and the generalization function.

## REFERENCES

Brayanov, J. B., Press, D. Z., and Smith, M. A. (2012). Motor memory is encoded as a gain-field combination of intrinsic and extrinsic action representations. *J. Neurosci.* 32, 14951–14965. doi: 10.1523/JNEUROSCI.1928-12.2012

Donchin, O., Francis, J. T., and Shadmehr, R. (2003). Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: theory and experiments in human motor control. *J. Neurosci.* 23, 9032–9045.

Emken, J. L., Benitez, R., Sideris, A., Bobrow, J. E., and Reinkensmeyer, D. J. (2007). Motor adaptation as a greedy optimization of error and effort. *J. Neurophysiol.* 97, 3997–4006. doi: 10.1152/jn.01095.2006

Fagg, A. H., Shah, A., and Barto, A. G. (2002). A computational model of muscle recruitment for wrist movements. *J. Neurophysiol.* 23, 1107–1112. doi: 10.1152/jn.00621.2002

Harris, C. M., and Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature* 394, 780–784. doi: 10.1038/29528

Hirashima, M., and Nozaki, D. (2012). Learning with slight forgetting optimizes sensorimotor transformation in redundant motor systems. *PLoS Comput. Biol.* 8:e1002590. doi: 10.1371/journal.pcbi.1002590

Huang, H. J., and Ahmed, A. A. (2014). Reductions in muscle coactivation and metabolic cost during visuomotor adaptation. *J. Neurophysiol.* 112, 2264–2274. doi: 10.1152/jn.00014.2014

Huang, H. J., Kram, R., and Ahmed, A. A. (2012). Reduction of metabolic cost during motor learning of arm reaching dynamics. *J. Neurosci.* 32, 2182–2190. doi: 10.1523/JNEUROSCI.4003-11.2012

Ingram, J. N., Flanagan, J. R., and Wolpert, D. M. (2013). Context-dependent decay of motor memories during skill acquisition. *Curr. Biol.* 88, 3348–3358. doi: 10.1016/j.cub.2013.04.079

Ingram, J. N., Howard, I. S., Flanagan, J. R., and Wolpert, D. M. (2011). A single-rate context-dependent learning process underlies rapid adaptation to familiar object dynamics. *PLoS Comput. Biol.* 7:e1002196. doi: 10.1371/journal.pcbi.1002196

Jones, K. E., Hamilton, A. F., and Wolpert, D. M. (2002). Sources of signal-dependent noise during isometric force production. *J. Neurophysiol.* 88, 1533–1544. doi: 10.1152/jn.00985.2001

Krakauer, J. W., Ghilardi, M.-F., and Ghez, C. (1999). Independent learning of internal models for kinematic and dynamic control of reaching. *Nat. Neurosci.* 2, 1026–1031. doi: 10.1038/14826

Lee, J. Y., and Schweighofer, N. (2009). Dual adaptation supports a parallel architecture of motor memory. *J. Neurosci.* 29, 10396–10404. doi: 10.1523/JNEUROSCI.1294-09.2009

Scheidt, R. A., Dingwell, J. B., and Mussa-Ivaldi, F. A. (2001). Learning to move amid uncertainty. *J. Neurophysiol.* 86, 971–985.

Scheidt, R. A., Reinkensmeyer, D. J., Conditt, M. A., Zev Rymer, W., and Mussa-Ivaldi, F. A. (2000). Persistence of motor adaptation during constrained, multi-joint, arm movements. *J. Neurophysiol.* 84, 853–862.

Schmidt, R. A., and Lee, T. (1988). *Motor Control and Learning, 5th Edn.* Champaign, IL: Human Kinetics.

Shadmehr, R., and Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* 14(5 Pt 2), 3208–3224.

Smith, M. A., Ghazizadeh, A., and Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol.* 4:e179. doi: 10.1371/journal.pbio.0040179

Takiyama, K., Hirashima, M., and Nozaki, D. (2013). "Prospective error determines motor learning: a step towards a unified model of motor learning," in *The 23rd Annual Meeting of Neural Control of Movement* (Puerto Rico)

Takiyama, K., Katahira, K., and Okada, M. (2009). Exact inference in discontinuous firing rate estimation using belief propagation. *J. Phys. Soc. Jpn.* 78:4003. doi: 10.1143/JPSJ.78.064003

Takiyama, K., and Okada, M. (2011). Detection of hidden structures in nonstationary spike trains. *Neural Comput.* 23, 1205–1233. doi: 10.1162/NECO/a/00109

Takiyama, K., and Okada, M. (2012a). Maximization of learning speed in the motor cortex due to neuronal redundancy. *PLoS Comput. Biol.* 8:e1002348. doi: 10.1371/journal.pcbi.1002348

Takiyama, K., and Okada, M. (2012b). Recovery in stroke rehabilitation through the rotation of preferred directions induced by bimanual movements: a computational study. *PLoS ONE* 7:e37594. doi: 10.1371/journal.pone.0037594

Tanaka, H., Krakauer, J. W., and Sejnowski, T. J. (2012). Generalization and multirate models of motor adaptation. *Neural Comput.* 24, 939–966. doi: 10.1162/NECO/a/00262

Tanaka, H., Sejnowski, T. J., and Krakauer, J. W. (2009). Adaptation to visuomotor rotation through interaction between posterior parietal and motor cortical areas. *J. Neurophysiol.* 102, 2921–2932. doi: 10.1152/jn.90834.2008

Taylor, J. A., Hieber, L. L., and Ivry, R. B. (2012). Feedback-dependent generalization. *J. Neurophysiol.* 109, 202–215. doi: 10.1152/jn.00247.2012

Thoroughman, K. A., and Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature* 407, 742–747. doi: 10.1038/35037588

Yokoi, A., Hirashima, M., and Nozaki, D. (2011). Gain field encoding of the kinematics of both arms in the internal model enables flexible bimanual action. *J. Neurosci.* 31, 17058–17068. doi: 10.1523/JNEUROSCI.2982-11.2011

Yokoi, A., Hirashima, M., and Nozaki, D. (2014). Lateralized sensitivity of motor memories to the kinematics of the opposite arm reveals functional specialization during bimanual actions. *J. Neurosci.* 34, 9141–9151. doi: 10.1523/JNEUROSCI.2694-13.2014