# **RESEARCH ARTICLE** Why Professional Athletes Need a Prolonged Period of Warm-Up and Other Peculiarities of Human Motor Learning

Robert Ajemian<sup>1</sup>, Alessandro D'Ausilio<sup>2</sup>, Helene Moorman<sup>3</sup>, Emilio Bizzi<sup>1</sup>

<sup>1</sup>McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge. <sup>2</sup>Department of Psychology, University of Rome 'La Sapienza,' Italy. <sup>3</sup>Helen Wills Neuroscience Institute, University of California, Berkeley.

**ABSTRACT.** Professional athletes involved in sports that require the execution of fine motor skills must practice for a considerable length of time before competing in an event. Why is such practice necessary? Is it merely to warm-up the muscles, tendons, and ligaments, or does the athlete's sensorimotor network need to be constantly recalibrated? In this article, the authors present a point of view in which the human sensorimotor system is characterized by: (a) a high noise level and (b) a high learning rate at the synaptic level (which, because of the noise, does not equate to a high learning rate at the behavioral level). They argue that many heuristics of human skill learning, including the need for a prolonged period of warm-up in experts, follow from these assumptions.

*Keywords*: expert performance, motor learning, negative transfer, neural network, savings, sensorimotor system, variability of practice

mmediately prior to participating in a competitive event, professional athletes who play sports that require the execution of fine motor skills (e.g., tennis, golf, baseball) spend a considerable amount of time warming up to enhance their performance. Some of this time is spent stretching or doing various exercises to properly warm up their muscles, tendons, and ligaments. But most of this time is spent practicing the same highly skilled activity to be performed in the competition that follows. For example, professional tennis players spend at least an hour on the court hitting balls before a match. Golf professionals spend a significant amount of time on the putting green and the driving range prior to teeing off in competition. The ritual of pregame batting practice has long been a tradition of major league baseball. And so it goes with every such sport.

This pre-event practice embodies far more than an effort by conscientious professionals to gain every conceivable competitive edge. Rather, it is a necessary component of their ability to execute fine motor acts at the high level of performance for which their motor systems have been indelibly attuned through a lifetime of training. Without such practice, it is well known that the initial level of skilled performance drops considerably. A professional tennis player who has not taken a single warm-up stroke is likely to lose the first set played against an inferior opponent. A golf player who has not taken a single warm-up stroke is almost certain to perform poorly on the first couple of holes. A baseball player who has not taken batting practice is not likely to get a hit in the first at-bat.

This effect, observed in a range of motor skills across many levels of performance, is known as the *warm-up decrement*,

and it has been systematically studied since at least the late 19th century (for review, see Adams, 1961). Various psychologically oriented hypotheses have been proposed as potential explanations (Schmidt & Lee, 1999), including the idea of a *motor set*. None of these explanations has proven particularly successful, and the matter is still studied today, albeit sporadically, in the kinesiology and sports science communities. Interestingly, we know of no instance where this effect has been studied by motor neuroscientists. Why is that? Is the observation really so obvious or uninteresting, from the perspective of motor neuroscience, that it tells us little about motor learning? Is it not somewhat paradoxical that even though a fine motor skill, once learned, is never forgotten, a significant amount of warm-up time is required to bring it to the surface at a peak level of expression?

Certainly, robotic devices that perform fine motor skills require negligible warm-up time. Robotic ping-pong players have been developed (e.g., Andersson, 1988), and these devices, once turned on, require only a brief amount of warm-up time so that all parts achieve their appropriate operating conditions (temperature, state of lubrication, etc.). When that short interval has passed, the robot is ready to play pingpong at whatever level it is capable of playing ping-pong. It does not need the extended period of warm-up that an expert human ping-pong player needs to perform to its fullest potential. Does this difference between humans and robots in the performance of fine motor skills have any consequences?

There are two ways to explain an expert's need for extensive pre-event warm-up in the performance of a fine motor skill. One possibility is that humans, similar to machines, need to ensure that all of their important actuating parts have achieved the appropriate operating conditions. To this end, the muscles, tendons, and ligaments need to be properly stretched and exercised before an event to ensure peak performance. These human actuators happen to require a much longer period of warm-up than robot actuators in order to reach their peak operating state. If true, this explanation suggests that no great insight into the motor system can be gleaned from the warm-up decrement.

There is, however, another possible explanation. Perhaps the experts need the practice time to properly retune or recalibrate their motor systems. Thus, even if a skill is practiced on a daily basis, the motor system can become slightly miscalibrated from one day to the next so that it needs recalibration

Correspondence address: Robert Ajemian, McGovern Institute for Brain Research, Building 46, Room 6193, 77 Massachusetts Ave., Cambridge, MA 02139, USA. e-mail: ajemian@mit.edu

to enable peak performance. If this explanation is correct, then it leads to further questions. How is it that the system became miscalibrated in the first place, however slightly, if a distinct motor memory trace exists separating this skill from other skills? How is it that this same motor memory trace can persist at some level through years of disuse, if miscalibration can occur from one day to the next? These are questions that we seek to answer with a new framework for learning that has been posited for biological motor control (Ajemian, D'Ausilio, Moorman, & Bizzi, 2010).

#### **Theoretical Framework**

In a previous study (Ajemian et al., 2010), we made certain assumptions about the motor system. The first is that all motor skills are learned in a distributed fashion in an overarching sensorimotor network (i.e., it is useful to think of a single distributed network as learning everything). Further, at any point in time, the network will exhibit a single configuration, which dictates the extent to which any skill can be executed and which embodies the totality of all previous learning. These assumptions in no way preclude motor memory traces from becoming localized in the network, particularly as the level of expertise increases. But it does suggest that interference between skills always looms as a major problem because the motor system, being distributed, contains no hardwired mechanisms for explicitly segregating the neural resources devoted to a particular task. Of course, if different skills involve totally different sets of end effectors, then obviously a certain amount of segregation emerges for free simply from anatomical connectivity and topographic structure. Some models in the literature, such as the MOSAIC model (Wolpert & Kawato, 1998), begin with the assumption of explicit segregation of neural resources to different tasks without explaining how this segregation may emerge during the course of natural behavior. We feel that the problem of how such segregation arises is a central issue in understanding biological sensorimotor control and motor memory formation.

From our perspective, then, it is possible to envision the motor system as a huge sensorimotor network that is capable of mapping sensory inputs into motor outputs and selforganizing in response to error signals (see Figure 1). Further, the state of the system at any instant in time is fully characterized by the current network configuration, which can be represented as a single point in a high-dimensional abstract weight space. During learning, this point moves around in weight space in order to reduce the error of the practiced skills. More specifically, when an error arises, the system adjusts its internal configuration and therefore moves closer in weight space to a configuration that embodies the solution for the desired skill. To have learned multiple skills means that the network has arrived at a point in weight space that embodies the overlapping solutions for the different skills.



the input layer a pattern,  $\mathbf{P}$ , is entered and its values are propagated forward through the network via weighted connections to downstream nodes. One can imagine multiple intermediate or hidden layers, along with numerous feedback connections. The important point is that the activation values,  $\mathbf{Z}$ , at the output layer are compared with the target values,  $\mathbf{T}$ , to generate an error signal which is then fed back through the system to adjust the weights.

The goal of this network is to learn a set of sensorimotor skills, which in this framework can be represented as an arbitrary nonlinear functional map from inputs to outputs:

$$\langle \vec{X}^1, \vec{X}^2, \dots \vec{X}^p \rangle \mapsto \langle \vec{T}^1, \vec{T}^2, \dots \vec{T}^p \rangle$$

where the superscript corresponds to the number of the skill (P distinct skills are being learned in this example). So far, nothing has been done other than recapitulate the classical neural network framing for learning theory, for which there are many different specific instantiations (e.g., Grossberg, 1982; Haykin, 1999; Hertz, Krogh, & Palmer, 1991; Sutton & Barto, 1998).

This framework immediately presents the Stability-Plasticity dilemma (Carpenter & Grossberg, 1987). The network must be flexible so that new information can be stored. However, if the network is too flexible, then new learning may overwrite old learning, a phenomenon known as catastrophic interference in the neural network literature (French, 1992). Basically, the problem is that all skills in a distributed network share the same neural resources (weights and nodes), and so each time one skill is practiced, it slightly disrupts-or interferes with-the traces left by previous skills. Repeated practice of all skills concurrently would alleviate the interference problem by converging on a solution that accommodates all constraints (as happens with artificial neural networks in many data-mining applications). From a biological standpoint, however, concurrent practice of an entire motor repertoire is totally unrealistic. Skills are practiced when they

are practiced—one cannot practice every single old skill at the same time that one is learning a new skill—and the motor system must somehow cope with the interference that naturally arises.

A novel solution to this problem has been proposed in Ajemian and colleagues (2010). The basic idea is as follows. Biological signal processing, unlike electronic signal processing, is an extremely noisy process. This noise is detrimental from the standpoint of learning. If the learning rate is low relative to the noise, then learning becomes difficult, if not impossible, as network reorganization due to the noise swamps network reorganization due to the signal. However, suppose that the learning rate is boosted to unusually high levels. Although high learning rates usually lead to network instability, it was shown in Ajemian et al. (2010) that under certain conditions, high learning rates and high noise levels can coexist in a functioning network. This type of network exhibits very distinctive characteristics, the most important of which is that the network weights are always changing even when there is no behavioral error. This being said, the unusually high learning rates at the systems level do not translate into similarly high learning rates at the behavioral level because of the noise-that is, much of the network's weight adaptation is due to noise, as opposed to signal. In a certain sense, the high noise level and the high learning rate neutralize each other, giving rise to a dynamic system with an effectively lower learning rate.

Although the plasticity in the Stability-Plasticity dilemma is clearly addressed in this framework, what about stability? If network synapses are constantly changing, then does the practice/performance of every single skill disrupt the memory for every other skill that has overlapping sensorimotor requirements? The answer is a qualified yes. In general, skills with overlapping sensorimotor requirements interfere. However, with excessive practice, it is possible for skills to reach a state of noninterference, which is mathematically represented by the concept of network orthogonality (Ajemian et al., 2010). Orthogonal states cannot, in general, be obtained in normal networks because the learning rate and noise levels are not high enough. Basically, the high learning rate and high noise level together give the network the "energy" it needs to explore its configuration space and find orthogonal solutions. Once the skills in a skill set have become orthogonalized, then future practice/performance of those skills will not lead to interference.

The overarching point we wish to make in this article is that the kinesiology and sports science communities have, for the last several decades, discovered a variety of heuristics for how humans best learn skills (for a review, see Schmidt & Lee, 1999). Some examples are the benefits of interleaved practice effect, accelerated relearning, negative transfer, and the variability of practice effect. Different high-level theories have been proposed for explaining these effects, such as motor programs and motor schema. We are instead proposing that many of the distinctive characteristics of human skill learning result from two simple ideas: the human sensorimotor system (a) is very noisy and (b) exhibits a high learning rate. In support of this claim, networks that exhibit these two properties are simulated according to different practice regimens, and the results are presented subsequently as they pertain to three of the motor-skill learning heuristics (the benefits of interleaved practice effect is discussed in detail in Ajemian et al. [2010]).

#### Simulation Details

Details of the simulation studies are provided in Ajemian et al. (2010), although we briefly summarize the main points here. We use standard multilayer perceptrons (Rumelhart et al., 1986) for the simulations shown. Our networks differ from conventional multilayer perceptrons in that the learning rate and noise levels are much higher. Whereas a typical learning rate is 0.01, we use a learning rate of 0.3 (30 times higher) in simulations. Similarly, we add high levels of noise to our neural networks at all three levels of operation: nodal signal processing (15%), weight multiplication (15%), and weight change as a result of trial error (200%). These noise levels are at least an order of magnitude higher than what we have seen previously reported. As a control, we also run simulations using a conventional neural network, which exhibits a low learning rate (0.01) and no noise. In short, the simulations are run using two different types of neural networks: the network we propose, which exhibits a high learning rate and a high noise level and a conventional network, used as a control.

The basic task we simulate is a center-out reaching task, since this task serves as the basis for so many studies in neurophysiology and psychophysics. The input to the network is a target location and the output is a movement command. There are two input nodes for the input representation, and these nodes represent the Cartesian coordinates of the target location. There are eight output nodes cyclically arranged (the 1st node is a neighbor to the 2nd node and to the 8th node, etc.) for the motor output representation. Two hidden layers contain variable numbers of hidden nodes in different simulation runs. Error update is done using standard gradient descent, and the error is computed as a mean squared error between the actual output and the desired target output (Since the inputs and outputs are treated as numeric vectors without units, the error has no units.). For a given simulation run, all parameters are the same for the two different types of the network, so that the only differences are contained in the learning rate and the noise level. With these two different types of networks, we simulate, for a reaching task and perturbations of a reaching task, the following three well-known heuristics of motor skill learning: accelerated relearning, negative transfer, and the variability of practice effect.

## Results

#### **Accelerated Relearning**

Once a subject has adapted to a perturbation during, say, reaching movements, if the subject is exposed to that same

perturbation a second time, relearning is faster than the initial learning. Some have referred to this effect as savings (Krakauer & Shadmehr, 2006). This terminology is somewhat unfortunate because the analogy to Ebbinghaus' original use of the term in the context of syllable learning is, strictly speaking, incorrect. Whatever term is used to label this effect, its genesis remains straightforward from a qualitative standpoint: the motor system, having been exposed to the perturbation previously, somehow retains some trace of the learning and that residual trace facilitates more rapid relearning. The question at a more quantitative level is how to model the formation and retention of a motor memory trace. This problem has been recently addressed through a model that posits the existence of two distinct adaptive processes that operate on different timescales, one long and one short (Smith, Ghazizadeh, & Shadmehr, 2006). Certainly, the additional machinery of multiple adaptive processes operating on different timescales can be used to explain phenomena such as accelerated relearning. But why stop at two processes when more could be included? If a subject relearns a perturbation repeatedly and becomes correspondingly faster at relearning, do more processes need to be included on additional timescales in order to provide accurate fits? And what is the neurophysiological basis for these distinct processes?

We instead propose that the dynamics of accelerated relearning emerge naturally in distributed learning systems tasked with learning multiple similar skills in sequence. The basic idea is that the weights of the system are always moving to accommodate the most recently executed motor skill. When a perturbation has been learned, the network has migrated to a point in configuration space that enables a particular skill to be performed under that perturbation. From that moment onward, the network configuration will move in other directions depending on what other motor behaviors are practiced/performed, including the baseline skill without the perturbation if that is next in the sequence. However, since the network has had to arrive on at least one previous occasion at a solution point for the skill under perturbation, its configuration remains nearer to another such solution point than it would otherwise be. This is how the previous experience leaves a "trace," that is, it biases the system to locations that are closer to points on the same solution manifold, thereby allowing a new solution point to be more rapidly reached. This effect is seen with both types of neural networks that we simulate, so it is simply a property of distributed learning systems and requires no further explanation within that framework.

Figure 2 shows accelerated relearning for a network with a high learning rate and high noise level. The network is initially trained on the basic competence, in this case reaching to targets, such that the sensory input vectors of reach direction are transformed into the corresponding motor output vectors. Then a perturbation (in this case, a visuomotor rotation) is added for 2,500 trials, and performance under this perturbation is embodied by the initial learning curve shown with the dotted line. A washout block of 1,000 trials follows



der a perturbation is simulated. Even though the exact same overall competency is being learned with the exact same sequence of targets, note how the learning curve is significantly steeper the second time around. Eventually, the two curves asymptote at the same place.

(data not shown). Then, another learning block under the same perturbation is performed for 2,500 trials. The learning curve for this second block of 2,500 trials is aligned with the first learning block and presented as the solid line. It can be clearly seen that the second learning curve is steeper than the first, even though the sensorimotor network itself always uses the same learning rate. This finding is quite robust, as it arises across extensive variations in simulation parameters.

We emphasize that accelerated relearning is also seen in the case of a network with a low learning rate and no noise. Therefore, we cannot use this effect to distinguish between the two types of networks. Rather, these simulations demonstrate that accelerated relearning can be explained by the framework of a single neural network, with distributed representations, being asked to learn two similar skills in sequence. There is nothing new in this demonstration. Scientists familiar with neural networks have been aware of this effect for decades, because the idea of rapid relearning of a skill after perturbation is quite similar to rapid relearning after network "lesioning." In the latter case, a neural network is perturbed (the weights are changed, nodes are eliminated, etc.) after some function has been learned; in the former case, the function itself, not the network, is perturbed after it is initially learned. Rapid relearning after network lesioning has been studied extensively at least since the 1980s (e.g., Rumelhart et al., 1986). To our knowledge, the effect of savings has not been specifically explored in the neural network literature, and this circumstance derives not from any subtlety required in replicating the effect (since it is trivial), but from the peculiarity of what a perturbation means in the context of motor control. In partcular, a perturbation, such as a visuomotor perturbation or a dynamic perturbation, means that the same input (target direction) is mapped to different outputs (motor commands) depending on whether or not the perturbation is being applied. Hence, a neural network cannot simultaneously learn the reaching competency and its perturbation, since the problem is mathematically ill-posed.

### **Negative Transfer**

If an individual is an expert tennis player, then he or she should not play squash because the individual's tennis game will suffer due to the similarity of the skills. This effect, known as *negative transfer* (Schmidt & Lee, 1999), seems quite surprising, because clearly an expert tennis player is better at playing squash for the first time than an individual who has played neither racket sport. So, the direction of transfer is asymmetric: In the example, the individual's tennis expertise transfers to squash, but the squash practice interferes with the tennis expertise. Further, a novice tennis player who starts playing squash for the first time may actually improve in tennis. Once again, there is an asymmetry in transfer, as the practice of squash affects an expert versus a novice tennis player differently.

The model explains these effects through the concept of orthogonality. Basically, for someone to become an expert at tennis, that individual must have practiced tennis so much that the tennis strokes have become "orthogonalized" in the network-that is, each tennis stroke is orthogonal to each other tennis stroke and to every other skill in that individual's repertoire. Squash strokes are rather similar to tennis strokes, meaning that there is considerable sensorimotor overlap between the two. Thus, in a network whose weights are constantly changing, the practice of squash will displace the network from its finely tuned location in weight space where all the tennis strokes are orthogonal to each other and to everything else. In essence, the practice of squash disrupts the orthogonality of tennis in an expert. However, the expert tennis player will clearly be superior to the novice tennis player in terms of playing squash, because the network of the expert tennis player is already configured closer to a solution point for squash as a result of the sensorimotor overlap between the sports.

To demonstrate the effect of negative transfer, we first simulated a basic center-out reaching competence for 50,000 trials so that performance was at an expert level, and the movements in each reach direction were as orthogonal as possible to the movements in other reach directions. Then, we defined a similar competence (by similar, we mean mathematically similar in the functional sense, such that slightly different inputs are mapped to slightly different outputs so that the two mappings, although not identical, are nearby in function space). The network was subsequently trained on this similar competence for a variable number of trials, after which the network was once again presented with the original competence. Figure 3 shows a plot of the decrement in the basic reaching competence as a function of the amount of practice of the similar competence. The performance is an inverted exponential, meaning the decrease in performance

is initially rapid with respect to the amount of practice of the new competence, but at some point it levels off: no matter how much the new skill is practiced, performance at the old skill becomes no worse. This finding is robust, as it arises across variations in simulation parameters (including variations in the similar competence).

A conventional neural network with a low learning rate and no noise cannot replicate this effect. If, after the related competence is learned, the original competence is once again performed, the previous expertise returns almost immediately without an intervening performance decrement. There is no negative transfer in this case. The explanation lies again with the concept of orthogonality. In the case of networks with a high learning rate and high noise level, expertise requires that the network find a configuration in which the condition of orthogonality is satisfied. Through practice of a similar competence, the network can be displaced from this solution because of the high learning rate. However, in the case of the network with a low learning rate and no noise, expertise is not acquired by arriving at a network configuration for an orthogonal solution; rather, expertise is acquired simply by moving from the configuration of a good solution to an "expert" solution. Once this expert solution is attained, practice of a related competence is not sufficient to stop the system from immediately returning to an expert solution.

#### Variability of Practice Effect

According to this effect, the conditions of practice should be varied, rather than held constant, to maximize learning and retention during subsequent testing. Unfortunately, the term



**FIGURE 3.** Negative transfer. First, the reaching competency is learned to an expert level. Then a similar skill is learned. Learning this closely related skill does indeed disrupt performance of the original skill. Further, the more the new skill is practiced, the greater is the disruption, although the effect is modest in that performance only decreases a few percent.



**FIGURE 4.** A schematic of the variability of practice effect. *S* denotes the initial weight configuration of the sensorimotor system. *A* and  $A^*$  denote the configuration that needs to be reached for two related skills. Through points exist in an abstract, high-dimensional weight space such that each point is a network configuration. Through practice, the network takes steps moving in the direction of those skills. Note in this example that practicing *A* twice does not bring the network as close to the solution for that skill as practicing  $A^*$  once followed by *A*.

subsequent testing has two distinct usages in the literature. In one case, it is used to mean that the tested task is one of the practiced tasks. In a second case, the tested task is supposed to differ from the practiced tasks. For example, suppose that one group practices only task A<sub>0</sub>, whereas a second group practices task A0 together with the related tasks A1, A2, A3, and A4. The tested task could be task  $A_0$  or it could be some different task,  $A_5$ , that lies outside of the span of the initial tasks. In the latter case, the ability of the sensorimotor system to generalize across tasks is being explored. It is well known that neural networks are capable of better functional generalization when more examples of the function are provided during training. So the variability of practice effect holds trivially when the tested task differs from the practiced tasks. Subsequently, we focus on the case when the tested task is the same as the practiced task, and generalization is not an issue.

Consider the simplest case in which a skill is envisioned as a single instance of an input–output mapping. The variability of practice effect follows conceptually from the simple assumption that weight change is proportional to error in neural networks. Look at the simple schematic in Figure 4: **S** is the starting point in weight space, **A** is a point in weight space that accomplishes task **A**, and **A**<sup>\*</sup> is a point in weight space that accomplishes a close variant of task A. Mixing in both tasks can initially speed up the learning process by increasing the error—and hence the weight change—during each learning trial, without ever steering the learning process significantly off course. The variability of practice effect cannot, according to this explanation, persist once the network has arrived at or near **A**, so the theory would predict that it only applies during the initial stages of learning.

Now, we simulate the variability of practice effect, with a high learning rate and high noise network, as it pertains more broadly to a competency or set of skills. For example, suppose an individual wants to learn how to roller-skate on a specific smooth surface such as wood. According to this effect, if the individual is given a fixed amount of practice, it is best to devote some practice time to roller-skating on additional surfaces that vary along the rough-smooth continuum, even when all testing occurs on wood. To demonstrate this effect, we again simulated the learning of a reaching competency, which is a mapping of input vectors in sensory space to output vectors in motor space, across a circular workspace. However, there were two distinctly different learning schedules. In one condition (specific), the network was always presented with examples of the desired mapping. In the second condition (the variability of practice condition), the network was presented 60% of the time with examples from the desired mapping; the other 40% of the time the network was presented randomly with one of four similar mappings representing reaching under various mild perturbations. These mappings, when compared with the desired mapping, took slightly different input vectors to slightly different output vectors (so the mappings are nearby in function space). The simulation results are shown in Figure 5. To assess the progress that was being made toward learning the desired competency, we defined an overall performance error. This quantity was computed offline at each time step by taking the present network configuration and running all the desired input patterns through the network in a noiseless fashion to see what the average error would be across all instances of the competency.

According to the overall performance error metric, the network initially learns the desired reaching competency slightly quicker with the variable practice schedule. When one stops to think about it, this result is counterintuitive. The variable practice schedule, which involves practicing input-output mappings that are slightly different from the desired input-output mapping, initially enables quicker learning of the desired mapping than a practice schedule in which only the desired mapping is practiced. How can this be? How can learning something other than what you are trying to learn allow you to learn faster? The answer is found in Figure 5B, which plots the actual error on a trial-by-trial basis. The online trial-by-trial error is greater during the variable practice schedule since the network is being asked to learn more, that is, multiple similar mappings rather than a fixed mapping. This greater error can lead to faster learning, as long as the mappings are sufficiently similar so that the network configuration is not being asked to move in radically different directions within weight space (see Figure 4). Here we see a tradeoff between learning speed and error size: faster learning can ultimately be achieved if higher short-term errors are tolerated.

It must be noted in Figure 5 that the variability of practice effect only holds up during the initial phase of learning. At some point (after about 180 trials) the specific practice schedule overtakes the variable practice schedule in performance level. According to the theory, such a crossover point must exist when the network moves to a point at or near the



desired competency. Thus, a prediction of the theory is that the variability of practice effect should diminish with time as skill performance improves. Finally, unlike the previous two effects, the variability of practice effect is not robust in our simulations, as it disappears for many parameter settings. The literature seems similarly ambiguous on the topic as the effect only seems to hold for some skills and not others.

This effect does not occur in conventional neural networks where the learning rate is low and there is no noise. The reason is straightforward: when the learning rate is low, order of practice effects (e.g., the *benefits of interleaved practice*) cannot occur at the level of a single or a few trials (Ajemian et al., 2010). They can only occur when the learning rate is high such that movements in configuration space are large relative to the geometry of the solutions themselves. (A mathematical analogue is the noncommutativity of rigid-body rotations in three-dimensional space: the order of rotations strongly influences the final configuration, except when the rotations are differential, in which case the order does not matter.)

#### **Decrement Warm-Up**

We began with the observation that professional athletes engaged in sports involving fine motor skills must extensively practice prior to performance in order to obtain the best possible performance results. The framework discussed offers a possible explanation. When an expert finishes practice the night before an event, the expert's sensorimotor system is finely calibrated at a point in weight space that enables a high level of performance. Suppose that the event is not scheduled until the next afternoon. The athlete has several waking hours (ignoring the effects of sleep) during which time other activities are performed, even if only rudimentary activities such as walking or playing a video game. Under the assumptions of a high noise level and a high learning rate, the weights of the athlete's sensorimotor network continue to change, thereby causing the network to become slightly miscalibrated (the state of orthogonality is perturbed). To restore the network to its previous level of expertise, practice is required. Note that for anyone other than a highly skilled athlete, the difference in performance with and without warm-up practice may not be as noticeable or not as much warm-up would be required to eliminate the performance differential. But at the highest skill levels where even slight performance decrements are noticeable and where the skills must be delicately refined until a state of orthogonality is achieved, even a few hours of normal network weight change may be sufficient to cause notable diminution of performance.

#### Discussion

We advocate the point of view that many of the distinguishing characteristics of human sensorimotor learning arise from two assumptions: our sensorimotor networks are very noisy, and they exhibit an unusually high learning rate. These two assumptions lead to unique network dynamics where the weights are constantly changing, even when there is little or no behavioral error. How can this framework be falsified or supported by further experiments? Although Ajemian et al. (2010) contains several such tests, here we propose a new one. If our theory is correct, then the reason why an expert athlete must warm up prior to performing has nothing (or at least very little) to do with simply warming up the muscles, tendons, and ligaments. Rather, the athlete is practicing to recalibrate a sensorimotor network that has become slightly miscalibrated. Therefore, no amount of generic warm-up of the relevant muscles, tendons, and ligaments should be able to substitute for actual practice of the motor skills themselves, which is necessary to restore these skills to a state of finely tuned orthogonality. In fact, the theory would even predict that a group experiencing an hour or more of warmup of the relevant muscles, tendons, and ligaments should perform no better than a group experiencing five minutes of such warm-up.

### REFERENCES

- Adams, J. A. (1961). The second facet of forgetting: A review of warmup decrement. *Psychological Bulletin*, 58, 257–273.
- Ajemian, R., D'Ausilio, A., Moorman, H., & Bizzi, E. (2010). The stable encoding of motor memories within noisy and redundant neural circuits. Manuscript submitted.
- Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37, 54–115.
- French, R. M. (1992). Semi-distributed representations and catastrophic forgetting in connectionist networks. *Connection Science*, 4, 365–377.

- Grossberg, S. (1982). *Studies of mind and brain*. Amsterdam: Kluwer Academic.
- Haykin, S. (1999). *Neural networks: A comprehensive foundation*. Upper Saddle River, NJ: Prentice Hall.
- Hertz, J. A., Krogh, A. S., & Palmer, R. G. (1991). Introduction to the theory of neural computation. Redwood City, CA: Addison-Wesley.
- Krakauer, J. W., & Shadmehr, R. (2006). Consolidation of motor memory. *Trends in Neuroscience*, 29, 58–64.
- Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. Volume I. Cambridge, MA: MIT Press.
- Schmidt, R. A., & Lee, T. D. (1999). *Motor control and learning*. Champaign, IL: Human Kinetics.
- Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biology*, 4(6), 3.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
- Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11, 1317–1329.

Submitted September 30, 2009 Accepted August 31, 2010