Assessing and Enhancing Complex Skill Learning with Virtual Environments: Basic Insights for Motor Rehabilitation

A Dissertation Presented
By
Meghan Elizabeth Huber
to
The Department of Bioengineering
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the field of
Bioengineering
Northeastern University
Boston, Massachusetts
June 10, 2016
ABSTRACT

Over recent decades, virtual reality (VR) and robotic technologies have demonstrated the potential to enhance physical therapy. Despite their advantages, clinical adoption of these technology-based systems have been slow due to limited evidence that they are more effective than traditional therapy. Currently, VR and robotic technologies are used to automate conventional therapy. Hence, the success of technology-driven rehabilitation relies on the efficacy of conventional therapy. This efficacy, in turn, is limited by our knowledge of motor learning and recovery.

The goal of this thesis is to address several open questions in motor learning necessary for enhancing the efficacy of VR and robotic rehabilitation. Currently these technologies are used to train simple movements, but VR and robotic systems also have to potential to assist the learning of more complex skills. Moreover, these technologies afford the ability to assess complex skill learning in a controlled fashion without needing to sacrifice task complexity. However, a basic understanding of skill learning is required for these systems to be used effectively. In order to identify motor learning principles that apply to relearning the motor skills relevant for everyday life, this research aims to better understand complex motor skill learning, in contrast to the more prevalent research on simple, highly-controlled laboratory movements.

Results from this research demonstrate that humans are sensitive to the redundancy of a task and learn strategies based on their variability and noise. Complementary experimental and computational results further reveal how the guidance methods used for shaping behavior should be adapted to the specific task. For tasks without redundancy, assistance should focus on reducing noise and variability. Contrastingly, assistance for
tasks with redundancy should guide subjects towards stable solutions where variability has minimal effect of task performance. Verbal instructions were also shown to affect complex skill learning depending on the control strategy and the type of task. Together, these results provide initial principles for guiding complex skill learning that is more akin to the learning of everyday skills. Furthermore, these results demonstrate how VR and robotic technologies can be harnessed to better understand motor learning and simultaneously identify new approaches for motor rehabilitation.
ACKNOWLEDGEMENTS

I am very fortunate to have been surrounded by such amazing family, friends, and colleagues over these past years, all of whom deserve my sincerest gratitude for supporting me through this journey.

First, I would like to thank my advisor, Dr. Dagmar Sternad. None of this research would have been possible without your guidance. I am eternally grateful for everything you have taught me. You have given me once-in-a-lifetime opportunities and instilled in me the qualities of a good, productive researcher that I will always remember.

I would also like to thank my committee members, Dr. Christopher Hasson, Dr. Miriam Leeser, and Dr. Sandra Shefelbine, for their insightful contributions to this research. All of you have also helped me grow and learn in ways beyond this dissertation. Science is far from an independent endeavor, and I must thank my mentors and colleagues who contributed to the research in this thesis, Adam Brown, Dena Guo, Dr. Nikita Kuznetsov, Dr. Allison Seitchik, Dr. Amee Seitz, and Dr. Stephen Harkins, for all your assistance, insights, and thoughtful discussions that led to this research. To the current and former members of The Action Lab, Zhaoran Zhang, Ian Zuzarte, Nick Korsantia, Fei Ye, Dr. Se-Woong Park, Dr. Pauline Maurice, and Dr. Anastasia Kyvelidou, I am fortunate to have worked with all of you, and most importantly, to have learned so much from you. To the undergraduate members of The Action Lab, Rebecca Cheng, Oliver Cervantes, Kathleen Owens, Hannah Tam, and Julie Wiedemann, it was such a pleasure to work with all of you. I appreciate all of the time and effort you put towards this research, and I cannot wait to see what great things you all accomplish in the future. To all of my friends and colleagues in the Autonomous Motion Department at the
Max Planck Institute for Intelligent Systems in Tuebingen, thank you for introducing me to the world of humanoid robotics and new ways to approach my own research.

I am also grateful to have met so many supportive and encouraging friends and colleagues through the Bioengineering department and the Graduate Women in Science and Engineering organization at Northeastern. I must also thank my knitting group at Circle of Stitches for the advice and mentoring you have given me over the past year, both academically and emotionally.

Finally, I would like to thank my family. To my brothers, Paul and Thomas, you two have always been there to brighten my life. Your visits have been some of my fondest memories in Boston, and I am so thankful to know that we will always have each other lean on. To the newest member of my family, Michael Keebler, I could never have done this without my best friend by my side. Thank you for always making me smile and laugh whenever I need it. I cannot wait to see what the rest of our life together has in store for us.

Most importantly, I would like to thank my parents, Pam and Paul. Everything I have accomplished in my life is thanks to you. You both dedicated your lives to opening doors and creating opportunities for me to succeed, and words cannot express how much I appreciate this. Thank you for always being supportive and there for me when I needed you. I will always remember your sage advice – “put my blinders on” and work hard.
# TABLE OF CONTENTS

1. Introduction................................................................................................................................. 1
   1.1 Prevalence of Motor Disorders and Need for Rehabilitation................................. 1
   1.2 Promise of Virtual Reality-Based and Robotic Rehabilitation......................... 4
   1.3 Importance of Understanding the Sensorimotor System for Rehabilitation....... 7
   1.4 Current Motor Learning Principles Used in Rehabilitation.............................. 8
   1.5 Need to Understand Complex Skill Learning for Improving Rehabilitation.... 11
   1.6 Thesis Overview .................................................................................................................... 13
   1.7 Summary of Thesis Contributions ...................................................................................... 22

2. Acquisition of Novel and Complex Motor Skills: Stable Solutions Where Intrinsic
   Noise Matters Less...................................................................................................................... 24
   2.1 Abstract ................................................................................................................................. 24
   2.2 Introduction: Acquisition of Skill and Experimental Control............................. 25
   2.3 A Methodological Approach to Understanding Skill Acquisition.................... 31
   2.4 Hypothesis: Exploit Redundancy and Seek Stable Solutions that Make Intrinsic
       Noise Matter Less.................................................................................................................. 33
   2.5 Paradigm 1: Skittles, a Discrete Throwing Skill......................................................... 33
   2.6 Paradigm 2: Rhythmically Bouncing a Ball: Humans Exploit Dynamically
       Stable Solutions.................................................................................................................... 40
   2.7 Paradigm 3: Complex Object Manipulation: Carrying a Cup of Coffee .......... 47
   2.8 Conclusions and Outlook ...................................................................................................... 52
   2.9 Figures .................................................................................................................................... 55

3. Skilled Motor Performance: Overcoming Limits in Temporal Precision from Neural
   Noise ........................................................................................................................................... 64
   3.1 Summary ............................................................................................................................... 64
   3.2 Results ................................................................................................................................... 65
   3.3 Discussion .............................................................................................................................. 73
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Sections</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>Methods</td>
<td></td>
<td>75</td>
</tr>
<tr>
<td>3.5</td>
<td>Figures and Tables</td>
<td></td>
<td>81</td>
</tr>
<tr>
<td>4.</td>
<td>Persistence of Reduced Neuromotor Noise in Long-Term Motor Skill Learning</td>
<td>4.1 Abstract, 4.2 Introduction, 4.3 Materials and Methods, 4.4 Results, 4.5 Discussion, 4.6 Conclusion, 4.7 Figures</td>
<td>85-109</td>
</tr>
<tr>
<td>5.</td>
<td>Implicit Guidance to Stable Performance in a Rhythmic Perceptual-Motor Skill</td>
<td>5.1 Abstract, 5.2 Introduction, 5.3 Experimental Methods and Design, 5.4 Results, 5.5 Discussion, 5.6 Conclusions and Outlook, 5.7 Figures</td>
<td>114-146</td>
</tr>
<tr>
<td>6.</td>
<td>The Effect of Stereotype Threat on Performance of a Rhythmic Motor Skill</td>
<td>6.1 Abstract, 6.2 Introduction, 6.3 Experiment 1, 6.4 Experiment 2, 6.5 Experiment 3, 6.6 General Discussion, 6.7 Figures</td>
<td>154-192</td>
</tr>
<tr>
<td>7.</td>
<td>Girls <em>Can</em> Play Ball: Stereotype Threat Reduces Variability in a Motor Skill</td>
<td></td>
<td>200</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>Abstract</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>7.2</td>
<td>Introduction</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>7.3</td>
<td>Baseline Experiment</td>
<td>205</td>
<td></td>
</tr>
<tr>
<td>7.4</td>
<td>Experiment with Stereotype Threat</td>
<td>209</td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>General Discussion</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td>7.6</td>
<td>Conclusions and Outlook</td>
<td>221</td>
<td></td>
</tr>
<tr>
<td>7.7</td>
<td>Figures and Tables</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>Discussion and Future Directions</td>
<td>228</td>
<td></td>
</tr>
<tr>
<td>8.2</td>
<td>Virtual and Robotic Technologies for Quantifying Motor Skill Performance</td>
<td>229</td>
<td></td>
</tr>
<tr>
<td>8.3</td>
<td>Differences Between Simple and Complex Motor Skill Learning</td>
<td>231</td>
<td></td>
</tr>
<tr>
<td>8.4</td>
<td>Psychological Influences on Motor Skill Learning</td>
<td>235</td>
<td></td>
</tr>
<tr>
<td>8.5</td>
<td>Future Directions for Improving Neurorehabilitation</td>
<td>237</td>
<td></td>
</tr>
<tr>
<td>8.6</td>
<td>Future of Virtual Reality and Robotic Rehabilitation</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Concluding Remarks</td>
<td>242</td>
<td></td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>244</td>
<td></td>
</tr>
<tr>
<td>Appendix A</td>
<td>Model for Rhythmic Ball Bouncing</td>
<td>294</td>
<td></td>
</tr>
<tr>
<td>Appendix B</td>
<td>Model for Rhythmic Ball Bouncing with Time-Dependent Manipulation</td>
<td>296</td>
<td></td>
</tr>
<tr>
<td>Appendix C</td>
<td>Model for Discrete Ball Bouncing</td>
<td>297</td>
<td></td>
</tr>
<tr>
<td>Appendix D</td>
<td>Statistical Comparisons of Baseline Participants with No-Threat (NT) Participants</td>
<td>299</td>
<td></td>
</tr>
<tr>
<td>Appendix E</td>
<td>Statistical Analyses of Baseline Participants with Stereotype Threat (ST) Participants</td>
<td>301</td>
<td></td>
</tr>
<tr>
<td>Appendix F</td>
<td>Validity and Reliability of Kinect Skeleton for Measuring Shoulder Joint Angles</td>
<td>303</td>
<td></td>
</tr>
<tr>
<td>F.1</td>
<td>Abstract</td>
<td>304</td>
<td></td>
</tr>
<tr>
<td>F.2</td>
<td>Introduction</td>
<td>304</td>
<td></td>
</tr>
<tr>
<td>F.3</td>
<td>Methods</td>
<td>307</td>
<td></td>
</tr>
<tr>
<td>F.4</td>
<td>Results</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td>Appendix</td>
<td>Title</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>F.5</td>
<td>Discussion</td>
<td>311</td>
<td></td>
</tr>
<tr>
<td>F.6</td>
<td>Conclusions</td>
<td>312</td>
<td></td>
</tr>
<tr>
<td>F.7</td>
<td>Figures And Tables</td>
<td>313</td>
<td></td>
</tr>
<tr>
<td>G.1</td>
<td>Abstract</td>
<td>315</td>
<td></td>
</tr>
<tr>
<td>G.2</td>
<td>Introduction</td>
<td>316</td>
<td></td>
</tr>
<tr>
<td>G.3</td>
<td>Methods</td>
<td>318</td>
<td></td>
</tr>
<tr>
<td>G.4</td>
<td>Results</td>
<td>321</td>
<td></td>
</tr>
<tr>
<td>G.5</td>
<td>Discussion</td>
<td>322</td>
<td></td>
</tr>
<tr>
<td>G.6</td>
<td>Conclusions</td>
<td>323</td>
<td></td>
</tr>
<tr>
<td>G.7</td>
<td>Figures and Tables</td>
<td>324</td>
<td></td>
</tr>
</tbody>
</table>
1. Introduction

We often marvel at the exceptional motor skills of elite athletes, dancers, and artists. However, we usually take for granted just how remarkable our own “ordinary” motor skills are. It is often not until our movements are impaired that we fully appreciate our motor capabilities, ranging from grasping a cup to throwing or catching. Even when we experience only temporary impairments from straining a muscle or fracturing a bone, basic tasks that could once be completed with ease suddenly require tremendous effort. Thus, when such motor impairments may become more permanent or even deteriorate over time, they can have devastating effects on a person’s quality of life. However humans are resilient, both physiologically and psychologically, and have a huge potential for recovery. Thus, assistance to reinstate motor function – rehabilitation – is extremely important as it gives patients the opportunity to recover as much of their ability to function independently in life as possible.

1.1 Prevalence of Motor Disorders and Need for Rehabilitation

Performance and learning of motor skills entail multiple processes of perception, motivation, cognition, and action. Hence, numerous disorders or injuries affecting the nervous system cause some degree of motor impairment. Often these motor impairments significantly impact a person’s ability to function independently in daily life. To demonstrate the pervasive presence of such motor disorders and the urgent need for effective rehabilitation, I will briefly review some prevalent causes of neuromotor impairment.
1.1.1 Developmental Disorders

In the US, approximately one out of every six children between the ages of 3 and 17 has a developmental disorder (Boyle et al., 2011). While the term developmental disorder constitutes a broad number of conditions involving different types of serious impairments, the most common childhood disorders are typified by either primary or secondary impairments in motor function. Cerebral palsy (CP) and developmental coordination disorder (DCD) are two of the most common primary motor disorders in children. CP affects between approximately every 1 in 323 children in the US (Christensen et al., 2016), and DCD affects 1.4 to 19% of school-aged children (Kadesjö & Gillberg, 1999; Lingam, Hunt, Golding, Jongmans, & Emond, 2009; Tsiotra et al., 2006). Autism spectrum disorders (ASD) are another type of commonly observed development disorders, affecting one out of every 68 children in the US (Christensen et al., 2016). Though ASD is primarily characterized by impairments in communication, social interaction, restricted interests and repetitive behavior (American Psychiatric Association, 2013), individuals with ASD often exhibit significant motor dysfunction, including impairments in gait and balance, arm function, and movement planning (Fournier, Hass, Naik, Lodha, & Cauraugh, 2010).

The motor symptoms of these developmental disorders are present in early childhood, and while they do not usually get worse over time, they persist throughout the individual’s lifetime (Bax et al., 2005). Thus, a lifetime of therapy is required to help with functional challenges resulting from this disorder, which incurs significant costs for the family and the health care system. For example, the medical costs for a child with CP is approximately 10 times higher than for a typically developing child (Kancherla, Amendah, Grosse, Yeargin-Allsopp, & Van Naarden Braun, 2012).
1.1.2 Neurodegenerative Diseases

Neurodegenerative diseases result from progressive degeneration and/or death of neurons and become more prevalent with age. One of the most common neurodegenerative disorders that causes motor impairments is Parkinson’s Disease (PD), which affects approximately 1.5% of the population over 70 years old (Pringsheim, Jette, Frolkis, & Steeves, 2014). Motor neurons are also susceptible to degeneration as seen in Amyotrophic Lateral Sclerosis (ALS) and progressive muscular atrophy. Research on pharmacological interventions and brain stimulation actively aims to slow, stop, and potentially reverse the processes of neurodegeneration. While these approaches are still in their infancy, motor rehabilitation and occupational therapy is provided to patients to increase functional independence as best as possible.

1.1.3 Acquired Brain Injuries

An acquired brain injury consists of any damage to the brain that occurred after birth that is not related to a congenital or a degenerative disease (Marshall et al., 2007). Examples of such acquired injuries include infection, tumors, toxin exposure, traumatic brain injury, and – the most common injury - stroke. While the impairments from an acquired brain injury are sometimes only temporary, they often result in permanent disability. In fact, stroke is currently the leading cause of adult disability worldwide. If regaining motor function is physiologically possible, motor rehabilitation is often required for the remainder of one’s life. Such rehabilitation can be extremely expensive and stroke alone results in health care costs of over $73 billion a year in the US (Towfighi & Saver, 2011).
1.1.4 Need for Long-Term Motor Rehabilitation

Motor dysfunction can arise at any point in life, at birth, after an injury, or with old age. Regardless of when the symptoms begin, they typically last throughout the individual’s lifetime. Hence, patients with motor disorders typically require long-term rehabilitation in order to maintain or increase their quality of life. Yet, obtaining the necessary amount of continual therapy can be challenging as clinical visits are often time consuming and expensive. Further, patients may not always have access to the resources needed for meaningful recovery. Thus, the development of efficient treatment in the clinic and in the home has been priority in the medical community. Although there is a heavy reliance on in-home rehabilitation, quantifying the true efficacy of in-home therapy is challenging due to the lack of monitoring capabilities. Hence, clinicians desire the ability to track progress of an in-home therapy regime. Because of these needs, virtual reality and robotic technologies have emerged as a promising approach to enhance conventional motor rehabilitation.

1.2 Promise of Virtual Reality-Based and Robotic Rehabilitation

Over the past two decades, virtual reality-based (VR) and robotic rehabilitation systems have been developed to (re)train the control of gait (Mirelman, Bonato, & Deutsch, 2009; Roy et al., 2009), balance (Kim, Jang, Kim, Jung, & You, 2009; Lange et al., 2011), upper arm (Burdea, Cioi, Martin, Fensterheim, & Holenski, 2010; Lo et al., 2010), and hand movements (Golomb et al., 2010; Huber et al., 2010; Morrow, Docan, Burdea, & Merians, 2006) to serve a multitude of different patient populations. These systems are advantageous because they can quantitatively measure performance, present real-time performance feedback, and enhance patient motivation (Burdea, 2003). They
are also very efficient at delivering therapy, both in the clinic and in the patient’s home. Compared to conventional methods of rehabilitation, VR and robotic technologies allow patients to execute a considerably larger number of trials within a given timeframe, which is often considered key for motor learning and recovery (Kleim & Jones, 2008; Lo et al., 2010; Merians et al., 2014; Wagner et al., 2011).

Despite these advantages, there is only limited evidence that current VR and robot-assisted interventions provide a clinical benefit to patients in terms of functional outcome (A. Henderson, Korner-Bitensky, & Levin, 2007; Merians et al., 2014; Norouzi-Gheidari, Archambault, & Fung, 2012; Saposnik & Levin, 2011). For example, the VA ROBOTICS study, one of the largest robotic rehabilitation studies to date, showed that intensive robot-assisted rehabilitation in a virtual environment indeed improved upper arm function in chronic stroke patients compared to usual care (Lo et al., 2010). While these results illustrate the promise of robotic therapy, the improvements gained from the robotic-assisted rehabilitation did not reach clinical significance when measured by the Fugl-Meyer (FM) scale, one of the most established outcome measures in stroke rehabilitation. On the FM scale, only improvements greater than 6 to 8 points are considered clinically meaningful for chronic stroke patients (Feys et al., 1998; Gladstone, Danells, & Black, 2002; Jørgensen et al., 1995b; Page, Fulk, & Boyne, 2012). After 3 months of robotic therapy, the average improvement in FM score was ~3.5 points, which was still shy of this minimal clinically important difference (Krebs & Hogan, 2012; Lo et al., 2010).

The second and equally important finding of this study was that standard rehabilitation had no effect on disability, impairment, or quality of life. In fact, this
revelation may explain why the robotic therapy did not yield greater improvement. If standard treatment does not improve motor function, it is not surprising that even with increased intensity, replicating the same ineffective treatment with VR and robotic technologies does not significantly improve motor ability. Perhaps it was the type of therapy delivered with the robotic systems, rather than the intensity that was the limiting factor of its effectiveness (Krebs & Hogan, 2012).

The results of several large-scale studies on gait training in stroke patients supported this disappointing result. The Lokomat robotic exoskeleton (Hocoma, Switzerland) was designed to automate the delivery of body-weight-supported treadmill training, a form of gait rehabilitation presumed to be more effective than conventional care (a variety of tasks aimed at increasing walking speed, endurance, postural stability, and symmetry). Critically examining the efficacy of the Lokomat revealed, however, that therapeutic intervention with the Lokomat was actually inferior compared to conventional care (Hidler et al., 2009; Hornby et al., 2008). A few years later, it was discovered that any body-weight-supported treadmill training was inferior to conventional care (Duncan et al., 2011).

At face value, these results appear to reinforce the most commonly voiced critique of VR and robotic interventions: despite being beneficial in terms of time and monetary costs, there is insufficient evidence for substantial clinical benefit over traditional therapy (Huber et al., 2010; Kitago & Krakauer, 2013; Wagner et al., 2011). However, a closer look at these results reveal that the limited efficacy of these interventions is not due to limitations of the technologies themselves, but rather due to our insufficient understanding of motor rehabilitation. In fact, the use of robotics and VR for
rehabilitation is quite promising. As demonstrated in the VA ROBOTICS study, while the benefits of robot-assisted therapy are not yet clinically significant, robotic intervention can lead to greater improvements compared to usual care, at least for upper limb rehabilitation (Lo et al., 2010). Our scant understanding of the behavioral and neural mechanisms behind motor recovery is currently limiting our effective use of VR or robotic technologies. There will undoubtedly come a propitious time when these technologies will play a major role in rehabilitation (Krebs & Hogan, 2012; Stein, 2012).

Presently, these technologies at least offer a useful platform to help address the knowledge gap that is currently slowing the development of more effective means of motor rehabilitation – our understanding of sensorimotor control and learning.

1.3 Importance of Understanding the Sensorimotor System for Rehabilitation

To further improve motor rehabilitation, we need a better understanding of (1) what patients need to learn in order to improve their motor function, and (2) how patients or even healthy individuals learn to improve motor function. Without this basic knowledge, deriving successful, evidence-based interventions to enhance motor learning, relearning and recovery for impaired patients remains problematic (Krebs & Hogan, 2012; M. F. Levin, Weiss, & Keshner, 2015).

All too frequently, we have only very limited information about how the brain is affected by the injury, disorder, or disease. For example, the neural underpinnings of motor symptoms associated with common childhood disorders – e.g., autism and developmental coordination disorder – are still unknown. And yet, even if we do know the neural substrate of the neurological disorder or injury, it usually remains unknown
how the brain dysfunction causes motor symptoms. Does the neural impairment affect the individual’s ability to sense, control, adapt and/or learn? Even though the appropriate type of therapy or assistance for a patient critically depends on the answer to this question, this is still an active line of research (Matthews, 2004; Wolpert & Flanagan, 2015). Hence, the present research aims to further shed light on how the nervous system learns to translate perceived sensory information into useful motor actions.

1.4 Current Motor Learning Principles Used in Rehabilitation

Although our current understanding of the neural control of movement is far from complete, there have been significant advancements in the field over the past century that have shaped our understanding of motor rehabilitation and recovery. Perhaps the greatest discovery from basic neuroscience to influence modern rehabilitation is the discovery of neuroplasticity, the fact that new neurons and neuronal connections can be formed even in the mature central nervous system. While the concept of neuroplasticity seems to be common knowledge nowadays, up until the 1960’s the prevailing presumption was that adult brain and spinal cord were “hard-wired” and incapable of producing new cells or connections that lead to changes in functional behavior (Fuchs & Flügge, 2014). Even Santiago Ramón y Cajal, the “father of neuroscience”, whose own work supported the notion of neuroplasticity, was initially not convinced of the idea (Fuchs & Flügge, 2014; Stahnisch & Nitsch, 2002). To quote Cajal, he believed that “in adult centers the nerve paths are something fixed, ended, immutable. Everything may die, nothing may be regenerated” (Cajal & May, 1991). Nowadays, increasing our understanding of the mechanisms of neuroplasticity is a prominent research area in neuroscience. One major objective of this work is to decipher how memories and learned behavior can persist
despite the ever-changing neural substrate (Ajemian, D’Ausilio, Moorman, & Bizzi, 2013; Matsuzaki, Honkura, Ellis-Davies, & Kasai, 2004; Yang, Pan, & Gan, 2009).

The discovery and acknowledgement of neuroplasticity has also had a significant influence on how motor rehabilitation is conducted today (H. S. Levin, 2006). It suggested that humans are indeed capable of regaining motor function at the neural level, although it is still debated whether that implies learning compensatory strategies or recovering lost motor control (Krakauer, 2015; M. F. Levin, 2011). Nonetheless, the discovery of neuroplasticity shifted the focus of intervention from muscle re-education to what we now know as neurorehabilitation (Wolf, 1983). It has also paved the way for methods that enhance motor learning and recovery at the neural level, such as brain stimulation techniques that can potentially assist the effects of behavioral interventions (Krakauer, 2015).

Results from extensive behavioral research on motor learning starting in the 1970s have also influenced current approaches to motor rehabilitation. The amount of practice, most prominently intensity and frequency, is naturally one major factor that determines one’s performance capabilities in a task (De Jong, 1957; Snoddy, 1926). Another important factor found to influence motor learning is feedback (Salmoni, Schmidt, & Walter, 1984; Schmidt & Lee, 2011). With the rise of digital technologies in the 1970s came the possibility of providing different forms of feedback beyond verbal information and demonstration alone. Hence, there has been a surge in interest of how different forms of added feedback could potentially expedite learning. Going beyond reinforcement as emphasized in the behavioristic literature, the rationale is that providing additional information increases one’s ability to correct performance errors for the next attempt.
Studies have found that either providing additional information about the individual’s end performance (knowledge of results) or motor behavior during task execution (knowledge of performance) can enhance learning (Salmoni et al., 1984; Schmidt & Lee, 2011).

Insights from this vast literature also indicate that it is critical how practice is structured for learning and retention. For example, practicing tasks in a blocked order (i.e., all trials of one task are completed before switching to the next task) results in faster learning, compared to practicing the tasks in a random order. However, random practice leads to better performance in retention tests due to a process referred to as contextual interference (Lee & Magill, 1983; Shea & Morgan, 1979). In fact, it is this body of literature that highlighted that learning, as compared to temporary performance improvements, is defined as the long-lasting changes, only seen in retention tests (Adams, 1987; Schmidt & Lee, 2011; Sternad, Huber, & Kuznetsov, 2014).

One example of how insights from neuroscience and behavioral research have previously informed the development of rehabilitation is constraint-induced movement therapy (CIMT). CIMT was designed for hemiparetic and hemiplegic patients in order to encourage the use of the affected limb by constraining the unaffected limb (Winstein et al., 2003). The premise is that intensive use of the affected limb induces neuroplasticity. To ensure that the learned behavior is correct, feedback is used during the rigorous training in order to guide or shape the affected arm towards the desired behavior. A large clinical trial found that this intervention led to persistent improvements in motor function in mildly to moderately impaired stroke patients 3 to 6 months post-stroke (Wolf et al., 2008).
Behavioral and theoretical insights into motor control have also played a role in designing robotic rehabilitation. For instance, in the VA ROBOTICS study, robotic technology only assisted patients as needed as opposed to passively guiding them through the exercises (Lo et al., 2009). The rationale for using this form of robot assistance came from neuroimaging results that suggest voluntary drive is critical for motor learning (Lotze, 2003). Moreover, the assistance provided was based on the minimum-jerk trajectory, characterized by a bell-shaped velocity profile as theoretically derived and experimentally verified as the optimal strategy for reaching (Flash & Hogan, 1985). For the robot-assisted therapy deviations from this trajectory were guided back to the nominal trajectory. By incorporating these insights into the training protocol, robotic therapy was able to yield a significant improvement on upper limb function compared to standard care. Unlike upper limb robotic therapy, lower limb robotic rehabilitation was not able to produce any beneficial effects, presumably because principles for control of locomotion were not sufficiently respected (Krebs & Hogan, 2012).

1.5 Need to Understand Complex Skill Learning for Improving Rehabilitation

To date, much of the recent research in motor learning has focused on the learning of rather simple tasks such as reaching to a point target or learning a finger tapping sequence. However, the principles of learning derived from these simple movement paradigms tend to be more akin to practical, rather than theoretical considerations. This is evident from the often contradictory results, such as the effect of type, frequency, and temporal delay of augmented feedback on performance (Salmoni et al., 1984; Schmidt & Lee, 2011).
In an attempt to go beyond the reporting of behavioral results and develop computational models of motor control, the robot-based experimental paradigm on reaching with adaptation to external perturbations emerged. Following a seminal study on adaptation to a robot-induced force field by Shadmehr and Mussa-Ivaldi (1994), a prolific line of research ensued that addressed questions on adaptation and generalization (Dizio & Lackner, 1995; Krakauer, Mazzoni, Ghazizadeh, Ravindran, & Shadmehr, 2006; Shadmehr & Moussavi, 2000), consolidation (M. Abe et al., 2011; Krakauer, Ghez, & Ghilardi, 2005), trial-by-trial learning (Donchin, Francis, & Shadmehr, 2003), and trajectory control (Dizio & Lackner, 1995; Wang & Sainburg, 2005), to name only a few. The paradigm employed simple horizontal reaching movements using a robotic manipulandum that afforded not only fine-grained measurement, but also delivery of perturbations via an external complex force field or a prismatic distortion of its path. When the perturbations were applied, subjects immediately committed large deviations from the desired path, but soon learned to adapt and compensate for this added perturbation. When the force field was removed, performance temporarily deteriorated as the compensation to the force field persisted, although subjects quickly returned to their prior behavior in the absence of the perturbation.

Examining adaptation and learning in these simple motor paradigms has shed light on basic sensorimotor control principles employing theoretical tools such as iterative learning, internal representations and Bayesian processes (Kawato, Furukawa, & Suzuki, 1987; Körding & Wolpert, 2004; Shadmehr & Mussa-Ivaldi, 1994). Yet, insights gleaned from these types of tasks are not guaranteed to translate into useful principles for rehabilitation, where patients need to relearn and retain more functional, complex skills.
Thus, the development of more effective rehabilitative interventions will benefit from a better understanding of how humans learn more complex motor skills.

1.6 Thesis Overview

The overall objective of this thesis is to contribute to our understanding of what and how humans acquire complex skills. However, studying complex skill learning is not without its challenges. Therefore, the first portion of this thesis describes an approach that quantitatively assesses complex skill learning. The second part of the thesis then uses this approach to examine learning of complex skills with the goal to identify principles that are relevant for the development of more effective rehabilitative interventions.

1.6.1 Approach to Assessing Complex Skill Acquisition with Experimental Control

In contrast to the theoretically grounded research on the simple experimental paradigms, research in sport science has frequently focused on exceedingly complex and specialized skills, such as landing a triple axel in figure skating (King, 2005) or pitching a 100-mph fastball (Fleisig, Bolt, Fortenbaugh, Wilk, & Andrews, 2011). However quantitative assessment of learning and control of these elite skills is limited. While it is possible to measure full body motion kinematics and kinetics during a fastball pitch, describing behavior alone is not sufficient to inform how skilled behavior is learned or controlled. This is especially true when the skill requires interaction with an external object, such as a ball or a slab of ice.

Ideally, research aimed at understanding motor learning principles for rehabilitation should use a melding of methods used to study simple and complex skills. On the one hand, simplified tasks permit precise assessment of performance variables, and thereby
afford scientifically stringent testing of hypotheses. On the other hand, tasks with higher complexity can provide us with insights into how humans learn functional skills similar to skills of daily living, although assessments tend to remain descriptive. Hence an approach that allows complex skill learning be understood with the same experimental control and theoretical rigor afforded by simple tasks is needed.

Chapter 2 proposes that virtual environments, when used judiciously, allow the experimenter to have considerable experimental control, without the need to sacrifice the task complexity. For instance, virtual environments allow tasks to have complex physics while still being tractable. After all, the experimenter defines the physics in the virtual workspace when rendering a dynamical object in a virtual environment. This virtual rendering has the advantage that it confines the task to exactly the variables and parameters that are analyzed. Hence, there are no uncontrolled aspects as would readily occur in any real experiment (Sternad et al., 2014). For example, in real-world dart throwing unrealistic assumptions must be made about the dynamics of the dart, like absence of drag or lateral forces, in order to simplify the analysis of human performance (Smeets, Frens, & Brenner, 2002). These inaccuracies create variability that can be mistaken as variability from the performer. This added variability ultimately makes it difficult to reliably extract control and learning strategies. In virtual environments, however, such assumptions do not need to be made, and major pitfalls can be avoided. Further and importantly, the virtual task can be manipulated at will to test hypotheses about the performance strategies.

Sternad and colleagues have previously utilized this approach to develop and assess learning on three virtual tasks: throwing a ball to a target, rhythmically bouncing a ball,
and carrying a cup of coffee. The novel contribution of this chapter is the formulation of the generalized approach for developing new experimental paradigms to study complex skill learning. Specifically, the detailed description of this approach informs the motor neuroscience community of methodological considerations for overcoming the existing challenges with studying complex motor control and learning.


1.6.2 Understanding and Enhancing Complex Skill Learning

The remainder of the thesis uses the approach described in *Chapter 2* to contribute to our understanding of how and what humans learn during complex skill acquisition. Purposely, this research aims to answer a set of specific questions about complex skill learning as relevant for the development of more effective rehabilitative interventions. These questions include:

- How does learning of simple tasks differ from learning complex tasks?
- How can feedback be delivered to shape behavior in the long term?
- How can verbal instruction affect and motivate complex skill learning?

1.6.2.1 How Does Learning of Simple Tasks Differ from Learning Complex Tasks?

Strategies used to learn simple tasks may differ from those used to learn more complex tasks as ubiquitous in daily life. If this is the case, then further study of complex skill learning is needed to better inform motor rehabilitation. Task redundancy (i.e. multiple solutions lead to task success) has been proposed as the minimal requirement for an experimental task to allow for the study of complex skill (Sternad et al., 2014). In
tasks with only one successful solution, reducing variability in motor execution is the only strategy available to improve task performance. When there is more than one possible task solution, however, subjects can employ a range of strategies to improve performance. For example, instead of reducing overall variability, subjects could shape their variability in a way such that it has minimal performance impact. Or they may seek solutions that have stability, such that explicit error corrections are less necessary.

Chapter 3 aims to identify differences between simple and complex skill learning by comparing learning strategies in two versions of a virtual throwing task. In one version of the task, subjects could learn to adopt noise-tolerant solutions. However in the second version of the task, it was impossible for subjects to utilize this strategy. Due to biomechanical and physical constraints of the human learner, this second version of the task was analogous to learning a simpler task without redundancy. In addition, long-term practice was examined to assess if and how learning strategies change during practice.

The novel contribution of this research is the demonstration that even though a task has redundancy, factors such as the topology of a task’s solution space along with biomechanical and physical constraints of the learner may limit certain learning strategies. This finding is important as it offers an explanation to discrepancies in the literature as to what variables are controlled during the complex throwing.

Additionally, this research is important for motor rehabilitation as it suggests that if a patient population has a specific motor impairment (i.e. reduced force generation, slowed movements, etc.), this constrains further limits learning. Particularly, a virtual throwing task was used as throwing a ball is a frequently used motor task for evaluating motor impairments in children (S. E. Henderson, Sugden, & Barnett, 2007). While rating
scales are employed to assess performance on these complex tasks, these assessments are unclear as to how or why these patients display particular deficits in these tasks. This study provides a method for assessing how specific motor impairments limit and inhibit learning in certain task configurations. Thus, this virtual task may become useful for a quantitative assessment of patients in the future.

- **Chapter 3:** Huber, M.E., Guo, D., & Sternad, D. (In submission). Skilled motor performance: Overcoming limits in temporal precision from neural noise. Formatted as a Report article for *Current Biology*.

### 1.6.2.2 How Can Feedback be Delivered to Shape Behavior in the Long Term?

According to received motor learning knowledge, feedback is one of the most critical variables that affect learning. Though it is common practice in motor rehabilitation to shape and improve behavior through quantitative and qualitative feedback, the underlying mechanisms leading to improved performance remain unclear. A second, less noted but equally critical question regarding the effect of feedback, is what happens when feedback is removed. Has the learner become dependent on feedback while practicing a skill or is the improved performance maintained when the feedback is removed? According to the guidance hypothesis, presenting feedback too often during practice can lead individuals to rely on that added information. As a result, performance deteriorates when the feedback is no longer provided (Salmoni et al., 1984; Schmidt, 1991). The goal of rehabilitation is that performance improvements achieved in the clinic will translate into daily life. Hence, deriving feedback paradigms that lead to improvements that are retained when the feedback is no longer available is crucial.

**Chapter 4** aims to develop an intervention to enhance learning of a virtual throwing task. First, this study determined how performance improved under real-world practice conditions (i.e. without any augmented information) over weeks of practice. Is this
improved performance achieved by increasing error correction or perhaps decreasing noise? Second, behavioral analysis and computational modeling were used to determine how reward feedback could be manipulated to reduce neuromotor noise, which is often assumed to be invariant. This study also examines whether improved performance obtained by manipulated reward feedback could be maintained after the manipulation was removed. Again, computational modeling explained why manipulating reward feedback could lead to a persistent reduction in neuromotor noise.

The first **novel contribution** of this research is the demonstration that the random fluctuations in overt motor behavior decrease with long-term practice. Generally, this noise component is assumed to be invariant in models of short-term motor learning. However, this study demonstrates that to account for the behavioral observations during long-term learning, the constant noise source used in iterative learning models must be replaced with a time-varying noise source.

The second novel contribution of this research is the result that manipulation of reward can be used to reduce noise even further. More importantly, this reduced noise persists for five days after the manipulation to reward is removed. This final observation has utmost importance in motor rehabilitation, as developing interventions that lead to persistent changes in behavior is a major priority. In addition, the behavioral and modeling results specifically inform how and why learning was enhanced with this manipulation.

Current research and virtual reality-based rehabilitation practices emphasize error- and reward-based learning, which is effective in many experimentally controlled tasks (M. Abe et al., 2011; Galea, Mallia, Rothwell, & Diedrichsen, 2015; Hinder, Tresilian,
Riek, & Carson, 2008; Shabbott & Sainburg, 2010; Wolpert, Diedrichsen, & Flanagan, 2011). However, activities of daily living are almost invariably redundant tasks that may require a different type of guidance. In complex tasks, reducing variability is not necessarily the dominant avenue for improving performance, as demonstrated in Chapter 3. For example, Sternad and colleagues have demonstrated that humans find solutions and shape their behavior such that variability and noise has minimal effect on task performance (Cohen & Sternad, 2009, 2012; K. Wei, Dijkstra, & Sternad, 2007). Instead, guidance should help the learner or patient identify the mapping between execution and task outcome and discover optimal solutions.

*Chapter 5* develops and tests a new means of accelerating learning of a novel perceptual-motor skill, rhythmically bouncing a ball on a racket. The theoretically-motivated method of guiding learners to the optimal solution in this rhythmic ball bouncing task required a manipulation of the task physics, which could not be implemented in the real world. Hence, the guidance method was implemented in a virtual environment. The specific intervention was informed by the theoretical analysis of the task. Specifically, previous research had shown that subjects seek dynamically stable solutions that are robust to small errors and noise. This theoretical insight was used as the basis for a manipulation intended to guide learners to find this solution. Note that this subtle intervention did not give explicit guidance and subjects were not aware of this. To assess whether this subtle intervention had effects on learning, a retention test was added to test if the learned behavior was maintained when the guidance was removed.

Prior studies attempting to guide learning in this same rhythmic ball bouncing task reported that their guidance methods either interfered with the learning (Marchal-Crespo,
Bannwart, Riener, & Vallery, 2014) or that the learned behavior did not persist upon removal of the guidance (Morice, Siegler, Bardy, & Warren, 2007). The novel contribution of this research is a theoretically-derived guidance method that can expedite learning in a rhythmic task. Just as in the previous study, this guidance method also achieved persistence of the learned behavior even upon removal of guidance. Moreover, this research demonstrates a novel approach for using virtual environments to guiding complex skill learning that goes beyond the commonly used explicit error- and reward-based approaches.


### 1.6.2.3 How Can Verbal Instruction Affect and Motivate Complex Skill Learning?

A multitude of variables can influence motor performance, as any coach or athlete can attest. Decades of research have shown that variables such as attention and practice schedule can impact motor performance (Schmidt & Lee, 2011). The influence of the subtleties of verbal instruction, however, has not received much attention (Krakauer, 2015). Whether it is training an elite athlete for the Olympics or retraining a patient to pick up a bag, verbal instruction from a coach or therapist is important for motor learning. While this instruction is typically used to relay requisite information about the task to the learner, it can also contain psychological and motivational cues that subtly but consistently influence how a person learns (Lee, Swinnen, & Serrien, 1994). It may be argued that the ability to use these cues effectively is what makes a successful coach or
therapist. Thus, an understanding of how and when to trigger these cues to enhance learning and maximize motor performance could be a useful tool for therapists, real or even “virtual” (Landin, 1994; Lange et al., 2012).

To examine the potential impact of increased motivation on motor performance, two studies assessed how a verbal instruction designed to evoke a stereotype might influence complex skill learning. Evoking a stereotype, such as differences in gender performance, is presumed to increase a person’s motivation to perform well in order to disprove the stereotype (Harkins, 2006; Jamieson & Harkins, 2011). While there are many possible ways to increase motivation through verbal instruction, our experiments employed a stereotype manipulation as it can be implemented in a controlled manner.

Chapter 6 specifically examines how increased motivation may impact crucial control strategies in a complex rhythmic ball bouncing task. Delivering the stereotype at different stages of learning assessed how the instruction’s effect depends on the learner’s skill level.

Chapter 7 examines how the same verbal instruction may have a different impact on a discrete version of the ball-bouncing task. This experiment was motivated by previous neuroimaging results suggesting that different motor strategies are used in discrete versus rhythmic performance (Hogan & Sternad, 2007; Howard, Ingram, & Wolpert, 2011; Ikegami, Hirashima, Taga, & Nozaki, 2010; Ronsse, Sternad, & Lefèvre, 2009; Schaal, Sternad, Osu, & Kawato, 2004; Sternad, Dean, & Schaal, 2000; Sternad et al., 2013). As the motor control demands are presumably different, this experiment tested stereotype effects on a discrete version of the same ball bouncing task. Detecting subtle effects in gross measures of motor performance is not always possible. Hence, both
studies used fine-grained, and potentially more sensitive measures of control, to identify any effects of increased motivation.

Previous studies assessing the effect of stereotype threat on motor performance only observed that evoking stereotype threat through instruction debilitated performance. The novel contribution of these two studies is finding that stereotype threat can in fact facilitate as well as debilitate performance in complex motor tasks. In addition, this research demonstrates that the effect of stereotype – debilitation or facilitation – depends on the control strategies that humans are using to perform the task. An understanding of how one can intentionally employ these motivational and instructional cues enhance learning and maximize motor performance is important for rehabilitation.


### 1.7 Summary of Thesis Contributions

The development of more effective rehabilitative interventions would benefit from a better understanding of how humans learn more complex motor skills. Hence the research in this thesis aims to contribute to our understanding of how and what humans learning during complex skill acquisition. To address these questions, a general approach to study complex skill learning using mathematical task models and virtual environments was formulated. This approach enables the study of complex skill learning in a hypothesis-driven manner with experimental control. The subsequent research studies use this approach and enrich our understanding of how and what humans learn during complex
skill learning. In summary, this research resulted in the following insights on skill learning:

- Complex tasks with redundancy afford more learning strategies than tasks without redundancy. However, the ability to utilize these additional strategies is constrained by the topology of the solution space of the task, as well as the physical and biomechanical constraints of the learner.
- Reduction in variability over long-term practice proceeds primarily through reduction in neuromotor noise.
- For tasks where reducing variability is critical for skill performance, manipulation of reward can further lower the noise component of motor variability. Moreover, this reduced noise persists for multiple days of practice after the manipulation is removed.
- For complex rhythmic skills, where exploiting dynamic stability is critical, manipulation of the dynamically stable solutions can expedite skill learning.
- Instruction that increases the individual’s motivation to perform the task well can differentially facilitate or debilitate skill performance depending on the correctness of the control strategy.

Additionally, a novel approach for guiding behavior in a complex skill that can only be implemented through a virtual environment is introduced. This approach reveals an avenue to utilize virtual environments that may inspire future directions for enhancing skill learning in both experimental and clinical settings.
2. Acquisition of Novel and Complex Motor Skills: Stable Solutions Where Intrinsic Noise Matters Less

This chapter outlines the approach developed by Sternad and colleagues to study the acquisition of complex motor skills in a controlled fashion by using virtual environments and robotic devices. This approach deviates from the more recent prevalent research focus on motor adaptation, where an existent movement pattern is adapted to external manipulations. It also contrasts with much of the previous work on motor learning that has focused on simple highly controlled experimental tasks that afford only a single correct solution. Instead, the approach described here allows for the quantitative study more complex tasks with more than one solution to understand the acquisition of novel skills. This is an essential process involving more strategic elements that is ubiquitous in life, enabling humans not only to perform extraordinary actions, such as skateboarding, but also to eat with knife and fork or drive a car. Through mathematical modeling and virtual rendering of experimental tasks, the methodological approach still affords hypothesis-driven and experimentally controlled research on more challenging skills.

I contributed to the literature review, formulation of the approach for studying complex skill learning, and presentation of the results.

2.1 Abstract

Most experimental paradigms in motor neuroscience have used relatively focal and experimentally constrained tasks to allow precise measurement and experimental control. Therefore, practice-induced improvements and learning has been confined to relatively simple changes or adaptations to external perturbations. Here, we propose an approach to
study more complex skills that are novel and require more extensive practice, leading to quantitative and qualitative changes in overt performance. Central to these skills is that they have extrinsic redundancy that allows exploration and exploitation of dynamic properties of the task. We hypothesize that in such skills, humans seek stable solutions that are robust to perturbations that make their intrinsic noise matter less. Three experimental paradigms exemplify our model-based and hypothesis-driven approach to skill acquisition: discrete throwing, rhythmic ball bouncing, and complex object manipulation. In skittles, a throwing skill, results show that actors are sensitive to the error tolerance afforded by the task. In ball bouncing, we show that subjects exploit the dynamic stability of the task, where small errors and noise self-stabilize without explicit corrections. In manipulating a “cup of coffee”, subjects learn to optimize the safety or energy margins and scale it to their intrinsic variability. This research presents new experimental paradigms that characterize the behavioral correlates of neuroplasticity in more complex skill acquisition. This fundamental work is a platform for future work to develop behavioral interventions for clinical applications.

2.2 Introduction: Acquisition of Skill and Experimental Control

While professional quarterbacks like Peyton Manning make passing a football look effortless, it takes a lifetime of disciplined practice for these elite athletes to reach such high level of skill. The principal responsibility of the quarterback in American football is to throw a forward pass to his teammate. This skill, however, is not as trivial as it may appear. To achieve a successful pass, he must accurately throw to a target that is moving, while he himself is in motion. But even before the pass is thrown, he must quickly decide which receiver will have the best chance to catch the ball, judging from the continuous
movements of the defensive team. He must also choose what style of throw and what initial conditions (i.e., position, velocity) are appropriate for where he aims the ball. This decision-making process cannot last more than a fraction of a second or he risks being tackled by opponents twice his size. This ability to choose from a vast – in fact infinite – set of actions is an integral part of many sport skills. Such complexity is what makes events like the Olympic Games fascinating, not only to spectators around the globe, but also to movement scientists, who want to unravel how humans achieve coordinated actions. How does the brain acquire and control such complex spatio-temporal skills?

It is somewhat sobering to then look into the laboratories and see how motor skill is typically studied. The vast majority of experiments have focused on highly simplified tasks, such as finger-to-thumb opposition (Karni et al., 1998), finger tapping (Wing & Kristofferson, 1973a), sequenced isometric contractions (Jae, Olafsdottir, Zatsiorsky, & Latash, 2005; Waters-Metenier, Husain, Wiestler, & Diedrichsen, 2014; Y.-H. Wu, Pazin, Zatsiorsky, & Latash, 2012), or reaching and pointing (Landi, Baguear, & Della-Maggiore, 2011; Shabbott & Sainburg, 2010; Shadmehr & Mussa-Ivaldi, 1994). Unfortunately, the requirements of experimental control, measurement, and quantification have stripped motor skills of most aspects that have made them interesting in the first place. Simplified tasks permit precise assessment of performance variables, show relatively consistent patterns across subjects, and thereby permit scientifically stringent testing of hypotheses. Such focal movements are also well suited for neuroimaging studies, such as fMRI, where the subject’s head must be still and more extensive movements distort the magnetic field. Yet, all these benefits come with the risk of oversimplification of task demands, to the point that findings from these tasks may no
longer generalize to more complex and realistic scenarios. These simple movements shed light on fundamental neural underpinnings, but they show little or no learning, as the movements are already part of the behavioral repertoire of the participants. How can learning and neuroplasticity be studied to look closer at the spectacular agility and complexity of movements in real life?

2.2.1 Motor Adaptation

One line of research that has been extremely successful and gone beyond the simple movement to study learning and neuroplasticity is the experimental paradigm on reaching with adaptation to external perturbations. Following a seminal study on adaptation to a force field by Shadmehr and Mussa-Ivaldi (1994), a prolific line of research ensued that addressed questions on adaptation and generalization (Criscimagna-Hemminger et al., 2003; Shadmehr & Moussavi, 2000), consolidation (Baraduc, Lang, Rothwell, & Wolpert, 2004), and trial-by-trial learning (Donchin et al., 2003; Thoroughman & Shadmehr, 2000), to name only a few research directions. The paradigm employed simple horizontal reaching movements that were perturbed via an external complex force field or a prismatic distortion of its path. It is important to point out that subjects initially perform straight-line reaches, as is intrinsic to their behavioral repertoire. When the perturbations were applied, subjects committed large deviations from the direct path, but soon learned to adapt and compensate, and their performance returned to the initial straight path. When the force field was subsequently removed, performance temporarily worsened, as the compensation to the force field persisted. These so-called after-effects indicated that the nervous system has learned something. In the absence of further perturbations, however, the return to the level prior to the experiment was relatively fast and persistent.
Can we consider these results to reveal mechanisms underlying the acquisition of skill? In our thinking, the answer is no, because at the end of the experiment, there is no net change in skill (Figure 2.1A). Even though there are after-effects and even though in a second encounter with the force field, adaptation may proceed faster (reflecting ‘savings’), these compensatory processes only re-establish, or maintain, previous behavior. The subject has not acquired any truly novel skill that she did not know before.

2.2.2 Skill Acquisition

Compare this motor adaptation to the scenario of learning how to throw a Frisbee. Here, the learner attempts a novel combination of arm, hand and body movements, with a novel object that has novel flight properties that she has never encountered before. With practice, she learns to release the Frisbee to lead to an accurate return, reflecting increased proficiency or skill (Figure 2.1B). Note that there is no imposed perturbation. The skill is new and persists after training, indicating retention. There is much anecdotal and some experimental evidence that once a motor skill is learned, it persists for a long time over years of no practice (Draganski et al., 2004; Nourrit, Delignières, Caillou, Deschamps, & Lauriot, 2003; Park, Dijkstra, & Sternad, 2013; Swift, 1910). While in motor adaptation there are also savings, it remains to be tested whether such savings persist for a similarly long time. If not, this would suggest that the temporary adaptation processes involve different forms of neuroplasticity than the acquisition of a novel skill. Hence, we conclude that the first desideratum for studying skill acquisition is that the movement task poses a novel challenge to the neuro-motor system.

2.2.3 Intrinsic and Extrinsic Redundancy

However, novel skills such as throwing a Frisbee are complex and difficult to study in the laboratory without some experimental reduction. Therefore the question is: what
are the minimal requirements for an experimental task to still allow for the study of skill, but in a controlled and hypothesis-driven fashion? We propose that the essence of skill is redundancy. As exemplified by the quarterback’s forward pass, a skill should allow for multiple solutions to achieve the task. Unlike in adapting a reaching movement in an external force field to restore the initial straight path, subjects should be able to explore different solutions to find the best one with practice. Such solutions may also differ between individuals.

Two types of redundancy can be distinguished: *intrinsic* to the organism and *extrinsic* in the task. The role of intrinsic redundancy has been widely recognized in motor control. Our complex neuromuscular system is hierarchical with an infinite number of degrees of freedom at any one level. The most frequently used example for redundancy is at the level of joints: pointing to an object with a multi-joint arm can be achieved with an infinite number of joint angle configurations (Bernstein, 1967). Equivalently, at the level of muscles, multiple muscle contraction patterns may lead to the same kinematic pattern (d’Avella, Saltiel, & Bizzi, 2003; Ting, 2007). This excess of degrees of freedom for any movement goal repeats itself at the level of motor units and further. However, the task remains unique: the finger needs to point to one location in space.

Extrinsic redundancy, in contrast, would allow pointing to a whole region in space. For example, grasping and bringing a cup of coffee can be achieved using multiple approach paths, multiple finger contact points, and multiple contact points with the mouth. Yet, the task is only successful if the person brings the cup to the mouth, drinks, and no coffee is spilled. There are infinitely many combinations of execution that lead to
success with zero error. This kind of task redundancy has received less attention in motor learning than the intrinsic redundancy. Returning to the quarterback example, not only can different throwing styles achieve the same target, also different throwing releases may lead to the same target. Extrinsic, or task, redundancy allows the actor to “decide” what combination of execution-relevant variables to use in pursuing the task goal. Note that “decision” does not always imply a conscious weighing of alternatives with high-level cortical involvement (Araujo, Davids, & Hristovski, 2006). Rather, the situation forces the quarterback to make extremely fast decisions, which may happen at much lower levels of the neural hierarchy. Hence, the second desideratum for studying skill and skill acquisition is that the task should include redundancy to allow for choices between many solutions that achieve a given task goal.

How can we capture this redundancy and study the acquisition of novel skills in the laboratory without losing experimental control and precise measurement? Unlike several decades ago, measurement of complex movements is no longer the problem, as both technology and analysis methods have significantly advanced. New camera systems and wearable sensors have outdated tedious digitizing of video recordings. In addition, virtual technology and programmable robotic devices have created a whole new arena for experimental manipulations and measurement. However, simply measuring trajectories of Peyton Manning’s throwing arm, even in truly ecological settings, does not tell us much about his skilled performance. Without experimental control and hypotheses, the multitude of data simply re-describes the performance. A hypothesis-driven approach to skill requires some simplification - but without “throwing the baby out with the bathwater”.
2.3 A Methodological Approach to Understanding Skill Acquisition

We developed an approach using mathematical modeling and virtual technology to study the acquisition of novel skills with extrinsic redundancy. In the following, we first describe our methodological steps and then present three examples following this approach: throwing a ball to a target, rhythmically bouncing a ball, and carrying a cup of coffee.

1) Select a skill. The first step is to choose a skill that represents some core aspects germane to many other tasks. In throwing, for example, the timing of release is one such core aspect. Carrying a cup filled with coffee exemplifies manual interaction with a complex object. Next, it is essential that the experimental task has a well-defined goal and yet allows for a variety of solutions to achieve this goal: Task redundancy should be center stage. For example, the quarterback throwing a ball to a receiver can use a variety of strategies to throw to the intended wide receiver. Further, the task should be novel and sufficiently challenging to require practice to achieve success. Improvement should be visible within one or few experimental session(s), but also require fine-tuning at a longer time scale.

2) Mathematically model the task. The next step is to model the relevant physics that govern the task. A model is a useful simplification of the real phenomenon because it formalizes the assumptions about the task and prunes away the irrelevant aspects of the real-life task. What is the simplest system that captures the challenge to manipulate a dynamically complex object? What exactly determines timing of release? One core element in the modeling stage is to distinguish between the execution variables and the result variables: The result variable(s) are defined by the task goal and capture the quality
of performance; this is typically an error measure. Execution variables are under control of the performer and fully determine the task result. The functional relationship between execution and result is the essence of the model.

3) **Formulate hypotheses.** Based on the mathematical model, the space of all possible solutions to the task can be derived. Note, the term “solution” is not confined to performances with zero error, but include those with non-zero error. As the model system is typically nonlinear, the space of solutions, or result space, is complex and reveals additional properties of solutions, such as risk or dynamic stability, as we explain below. Depending on the model, different mathematical tools can be used to derive predictions about stability or robustness to perturbations. Importantly, exact quantitative hypotheses can be formulated about which solutions have the greatest probability of success.

4) **Render the model task in a virtual environment.** Based on the explicit mathematical understanding, the task can be rendered in a virtual environment. The execution variables are the ones that the subject controls via an interface with the virtual system. The result variables can be precisely measured in the virtual world. For example, while the subject performs a throwing task, the arm trajectory is real and controls the ball release, but the ball and target are virtual. The virtual rendering has the advantage that it confines the task to exactly the variables and parameters that were analyzed. There are no uncontrolled aspects as would occur in a real experiment. Further, the execution and the result variables can be manipulated at will to test hypotheses about the performance strategies.

5) **Measure subjects’ performance and test hypotheses.** Subjects interact with the virtual physics of the task via a manipulandum. The variables measured from the
subjects’ movements should correspond to the execution variables in the model. The measured execution variables and the task result are then evaluated against the space of all solutions. Hypotheses about solutions derived from the model can be evaluated.

2.4 Hypothesis: Exploit Redundancy and Seek Stable Solutions that Make Intrinsic Noise Matter Less

Before detailing this approach, it is necessary to be more specific about the overall research aim. Given the intrinsic and extrinsic redundancy of the task and performer, we pursued the hypothesis that individuals improve their performance by seeking solutions that are robust with respect to perturbations that make their intrinsic noise matter less. Due to the intrinsic redundancy of the hierarchical and complex neuromuscular system, subjects always show variability or intrinsic noise, even when they want to repeat the same movement under fixed external conditions (Faisal, Selen, & Wolpert, 2008). In the presence of extrinsic redundancy with infinite equivalent task solutions, we hypothesize that the neuro-motor system chooses those solutions that are most tolerant to this ever-present noise. As highlighted above, tasks with redundancy provide options that subjects can explore and exploit. This overall hypothesis will be operationalized for each task on the basis of the model. The following review exemplifies how we pursued this approach in three different paradigms.

2.5 Paradigm 1: Skittles, a Discrete Throwing Skill

2.5.1 The Skill

This experimental paradigm was motivated by the British pub game “skittles”, which is a table version of the American game “tetherball”. The actor throws a ball that is
tethered to a post by a string like a pendulum; the goal is to hit a target skittle on the opposite side of the pole (Figure 2.2A). Accurate throwing requires a controlled hand trajectory that prepares the ball release at exactly the right position and velocity that sends the ball onto a trajectory that hits the target skittle. The task has redundancy as elaborated below and timing is one essential element of this skill.

2.5.2 The Model

To simplify the task, the movement of the ball was confined to two horizontal dimensions, eliminating the elevation due to the pendular excursion (Figure 2.2B). In the model, the ball is attached to two orthogonal, massless springs with its rest position at the center post. To execute a throw, the ball is deflected from its rest position by the hand—a virtual lever arm that moves in correspondence to the real arm movements of the participant (Figure 2.2B). Upon release, the ball traverses an elliptic path generated by the restoring forces of the two springs (Müller & Sternad, 2004b). The equations for ball position in the $x$- and $y$-directions at time $t$ are:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} x_p \\ y_p \end{pmatrix} \cos \omega t + \begin{pmatrix} \cos \phi_r & -\sin \phi_r \\ -\sin \phi_r & \cos \phi_r \end{pmatrix} \begin{pmatrix} l \cos \omega t \\ v_r/\omega t \end{pmatrix}$$

The frequency $\omega$ denotes the natural frequency of the springs, and $(x_p, y_p)$ denotes the pivot point and $l$ the length of the arm. For a given throw, two execution variables, angle $\phi_r$ and velocity $v_r$ of the virtual hand at the moment of release fully determine the ball trajectory.

Similar to the real skill, the actor’s goal is to throw the ball to hit the target skittle, without hitting the center post. The location of the target is defined in $(x, y)$-coordinates; the error is defined as the minimum distance between the ball trajectory and the center of the target (Figure 2.2B). Thus, the result variable is the error and is fully determined by...
the two execution variables. Importantly, there is more than one combination of angle $\phi_r$ and velocity $v_r$ that leads to zero error. Figure 2.2C illustrates this functional relationship: for each point in execution space, spanned by release angle and velocity, error is depicted by shades of gray. Lighter shades indicate smaller errors; black signifies that releases hit the center post, which is a penalty area in the experiment. The thin black lines denote the one-dimensional set of zero-error solutions, or solution manifold. The solution manifold can be analytically derived and is defined by the following equation:

$$\frac{v_r}{\omega} = \sqrt{\frac{|-(l \sin \phi_r - y_p)x_t + (l \cos \phi_r + x_p)y_t|}{(l + \cos \phi_r x_p + \sin \phi_r y_p)^2 - (\cos \phi_r x_t + \sin \phi_r y_t)^2}}$$

An interesting feature of this task is that the result space and the solution manifold depend on the target location in ways that would be hard to predict without the model (Figure 2.2C). For example, for target position at $x=0.6$ m and $y=1.5$ m, the result space is non-linear and U-shaped, as shown in the top left example (Cohen & Sternad, 2009). For the target position $x=0.05$ m, $y=1.055$ m, the solution manifold is a vertical line, such that only the angle of release determines the error, as shown in the top right example (Sternad, Abe, Hu, & Müller, 2011). In the two examples in the lower panels, the solution manifolds terminate on the penalty area, which makes those solutions risky. Importantly, the difficulty of the task is determined by the shape of the solution manifold and can vary significantly for different target locations.

2.5.3 Hypotheses

In principle, subjects can release the ball with any of the position-velocity combinations on the solution manifold. However, Figure 2.2D illustrates that this is not what subjects do: the two data sets represent 20 throws collected on the first day (blue)
and after 11 days of practice (yellow). Each data point on this figure corresponds to a throw and, as to be expected, the data show significant scatter, especially in the beginning. The two distributions also show that the subject improved, as the throws early in practice have visibly higher errors on average (in dark colored area). The throws late in practice are aligned with the solution manifold, with smaller errors. What drives this change? What solutions do humans prefer and why? The hypothesis is that individuals improve their performance by *seeking solutions that are robust with respect to perturbations to make their intrinsic noise matter less.*

The sample data illustrate how three conceptually different routes to performance improvement can be distinguished: 1) Subjects shift their solutions in execution space to those regions on the solution manifold, where small changes in the angle and velocity of release do not lead to large decreases in the magnitude of error. Wider bands of light gray on the right branch indicate that deviations from the solution manifold do not incur large errors. We refer to this as error *Tolerance.* 2) Subjects reduce the amplitude of their dispersion. We refer to this as *Noise Reduction.* 3) Subjects rotate or re-shape their distribution to align with the solution manifold. We refer to this as *Covariation.* All three routes channel variability to make their intrinsic noise matter less for the quality of the performance. The specific hypothesis is that learners first improve their performance by shifting their throws to error-tolerant solutions in result space. This first stage includes exploration of options. The second stage involves rotating, or covarying, the data with respect to the solution manifold. Having exploited these two routes, the final stage is to reduce the dispersion. This latter route corresponds to reducing intrinsic noise.
Quantitative measures of *Tolerance, Noise*, and *Covariation* were first introduced by Müller and Sternad (2003, 2004a, 2004b, 2009) and subsequently re-formulated as cost measures by Cohen and Sternad (2009). The *TNC-Cost* measures assess to what degree the result of a set of data could be improved, if one of these components were optimized. The numerical analysis method was detailed in Cohen and Sternad (2009). Note the similarity with the UCM-method (Scholz & Schöner, 1999), but Müller and Sternad (2009) also outlined some key conceptual differences between the methods. One critical methodological issue is that the UCM-method, as one of the many covariance-based methods, is more prone to coordinate transformations, although the TNC-method, in particular the Covariation-Cost, is not immune to this problem either (Sternad, Park, Müller, & Hogan, 2010).

### 2.5.4 Virtual Environment

To test the hypothesis, the mechanical model was rendered in a virtual setup, where subjects interacted with the model via a manipulandum. In the experiment, the actor grasped a ball attached to the distal end of a manipulandum and performed a forearm rotation to throw a virtual ball to a target in the virtual environment (Figure 2.3A). Initially, the subject closed a contact switch with her index finger; lifting the finger opened the switch and triggered the release of the virtual ball. A potentiometer continuously sampled the angular position of the manipulandum, and the movements of the arm were mapped onto the displayed virtual arm; angular velocity of the arm was estimated online. The ball traversed around the center post as determined by the angular position and angular velocity at the moment of release (Figure 2.3A).
2.5.5 Experimental Findings

Cohen and Sternad (2009) tested the hypothesis that learners improve their performance by first locating error-tolerant solutions in result space and then by channeling and reducing their intrinsic noise. In their study, 9 participants with average and 3 participants with extensive throwing experience practiced 3 blocks of 60 throws each day for 6 and 15 days, respectively. Each block of 60 throws was analyzed as a distribution and the three components Tolerance, Noise, and Covariation were quantified: Each data set was translated to obtain T-Cost, scaled for N-Cost, or oriented via recombination for C-Cost to find the optimal mean result for this component. Note that each transformation left all other aspects of the data invariant. The cost was then defined as the difference between the mean result, error, of the original data set and the optimally transformed data set.

The TNC-Cost analysis showed that both novices and expert throwers revealed similar patterns of change in T-, N-, and C-Costs over practice (Cohen & Sternad, 2009). Figure 2.3B shows all three costs for the expert throwers. T-Cost (red) decreased very early in practice followed by a considerably slower decrease in C-Cost (blue). N-Cost (green) decreased at the slowest rate and remained the highest cost to performance. The overall rank ordering of these costs was similar across participants and expertise levels. This pattern of results suggests that early in practice participants tried out different performance strategies and searched for the most error-tolerant region of the solution manifold. They then improved covariation to better align with the solution manifold. The stochastic Noise component was reduced only modestly, suggesting that intrinsic noise is relatively inaccessible to practice.
How does the actor achieve these changes? Particularly, how can covariation between position and velocity be achieved? Extending initial work by Müller & Loosch (1999), Cohen & Sternad (2012) tested the hypothesis that subjects learn to align their trajectory of the arm with the solution manifold. For this analysis, it is interesting to recognize that the execution space, spanned by position and velocity, is equivalent to state space. Hence, not only the release points, but also the continuous trajectory can be analyzed in this space (Figure 2.4A). The solution manifold corresponds to an ideal trajectory, which would achieve the desired result at every point. Evidently, the arm trajectory cannot exactly follow the solution manifold, but a trajectory that at least temporarily follows the solution manifold gives the subject a time window to release the ball that achieves a successful hit. Cohen and Sternad (2012) hypothesized that with practice, subjects shape their trajectories to travel along the solution manifold so that the ball can be released within this time window and achieve a successful hit. This strategy exploits the redundancy of the task and ameliorates the effects of intrinsic noise in timing.

The same participants’ data from the first study were analyzed. Each continuous arm trajectory was converted to a trajectory of error over time, assuming that every angle-velocity point on the trajectory was a release (its associated error was computed using the model equations). To measure the shaping of the arm trajectory with respect to the solution manifold, time-in-hit-zone was defined as the time that the trajectory spent below a given error threshold. Timing accuracy or timing error was defined as the time difference between actual and ideal release time. The ideal release time was the time when the arm trajectory crossed the solution manifold. Cohen and Sternad (2012) found
that with extended practice, performance improved due to continued shaping of the trajectory (Figure 2.4A). Interestingly, timing error decreased first but then asymptoted at 9 ms at approximately Day 6, while time-in-hit-zone continued to improve (Figure 2.4B). This suggests two interpretations: When timing error reached a plateau, performance improvement could only be achieved by aligning the trajectory with the solution manifold. Alternatively, subjects did not need to further improve their timing accuracy as their trajectory was in the hit zone and made any further improvement in timing superfluous.

2.5.6 Interim Conclusions

Analysis of the throwing skill exemplified our approach to skill acquisition using a discrete task with redundancy. The results showed that redundancy allows the performers to choose their performance strategy to minimize the potentially detrimental effects of intrinsic noise on performance. These findings support the overall hypothesis that when performing a new skill, humans exploit the redundancy of the task and find the error-tolerant regions to enhance their performance.

2.6 Paradigm 2: Rhythmically Bouncing a Ball, Humans Exploit Dynamically Stable Solutions

2.6.1 The Skill

Rhythmically bouncing a ball on a racket is a seemingly simple and playful task. Yet a closer look reveals that it requires a high level of perceptually-guided coordination to repeatedly intercept the ball to hit a target amplitude, i.e. perform in a rhythmic fashion (Figure 2.5A). As in skittles, success is determined at one critical moment when the racket hits the ball, as this impact fully determines the trajectory of the ball. Hence, a core
feature of this task is the control of collisions, which is germane to numerous other behaviors, ranging from running to clapping, and numerous sport skills, such as volleyball. One key difference to skittles, however, is that these impacts are performed in continuous rhythmic fashion, and errors from one bounce propagate to the next. Via the repeated interactions with the ball, the actor becomes part of a continuous dynamical system.

2.6.2 The Model

The model for this task is a well-studied nonlinear dynamical system, originally developed for a particle bouncing on a vibrating surface (Guckenheimer & Holmes, 1983; Tufillaro, Abbott, & Reilly, 1992). The simple model consists of a planar surface moving sinusoidally up and down, and a ball or point mass that impacts the surface with instantaneous contact (Figure 2.5B). The vertical position of the ball \( x_b \) between the \( k \)th and the \( k+1 \)th racket-ball impact follows ballistic flight:

\[
x_b(t) = x_r(t_k) + v_b^+ (t - t_k) - g/2(t - t_k)^2
\]

where \( t_k \) is the time of the \( k \)th ball-racket impact, \( x_r \) is racket position, \( v_b^+ \) is the ball velocity just after impact, and \( g \) is the acceleration due to gravity (9.81 m/s\(^2\)). With the assumption of instantaneous impact, the ball velocity just after impact \( v_b^+ \) is determined by:

\[
v_b^+ = (1 + \alpha)v_r^- - \alpha v_b^-
\]

where \( v_b^- \) and \( v_r^- \) are the ball and racket velocities just before impact, and the energy loss at the collision is governed by the coefficient of restitution \( \alpha \). The maximum height of the ball between time \( t_k \) and \( t_{k+1} \) depends on the ball and racket velocities just before impact and the position at impact \( x_r \), as shown by the following equation:
\[ \max_{t_k \leq t \leq t_{k+1}} x_b(t) = x_r(t_k) + (((1 + \alpha) v_r^- - \alpha v_p^-)(t - t_k))^2 / 2g \]

The task goal is to bounce the ball to a target height, and the error is defined as the difference between the maximum height of the ball and the target height (Figure 2.5B). Even in this simplified form, the task has redundancy, as the result variable error is determined by three execution variables: \( v_r^- \) and \( v_p^- \) and \( x_r \). Figure 2.5C shows execution space with the solution manifold, i.e. the planar surface that represents all solutions leading to zero height error. The blue and yellow data points are two data sets from early and late in practice, respectively; each data point corresponds to one ball-racket contact. As to be expected, the early data show a lot of scatter, while the late practice data cluster around the solution manifold.

### 2.6.3 Hypotheses

The representation of bounces in execution space affords analysis of the data using the TNC-method. As in the two-dimensional version in skittles, the data can be transformed to estimate the contributions of Tolerance, Noise, and Covariation. The same hypothesis as in skittles can be pursued in ball bouncing. However, an important difference to skittles is that only racket position and velocity at impact are under the actor’s direct control; the third execution variable, ball velocity, is determined by the previous contact (which in turn is determined by the previous contact, etc). Therefore, continuous rhythmic control of the ball depends on the preceding contacts, which adds an additional challenge to the controller. As mentioned above, the task is a dynamical system that lends itself to different analyses and predictions.

To model the racket and ball system as a continuous dynamical system, the racket movements need to be specified. To this end, the model assumes simple sinusoidal
motion of the racket. With this assumption, racket position and velocity at impact collapse into a single state variable, racket phase, or acceleration at impact. Further, the continuous system can be discretized at the ball racket contact, where ball and racket position are identical. A discrete map can be derived based on two state variables, the ball velocity just after impact $v_k^+$ and the racket phase at impact $\theta_k$:

$$v_{k+1}^+ = (1 + \alpha)A\omega \cos \theta_{k+1} - \alpha v_k^+ + g\alpha(\theta_{k+1} - \theta_k)/\omega$$

$$0 = A\omega^2(\sin \theta_k - \sin \theta_{k+1}) + v_k^+ \omega(\theta_{k+1} - \theta_k) - g/2(\theta_{k+1} - \theta_k)^2$$

where $A$ and $\omega$ are the amplitude and frequency of the sinusoidal racket movements (Dijkstra, Katsumata, De Rugy, & Sternad, 2004; Schaal, Sternad, & Atkeson, 1996).

This ball-racket system is a nonlinear system that displays dynamic stability and, despite its simplicity, shows the complex dynamics of a period-doubling route to chaos (Guckenheimer & Holmes, 1983; May, 1976; Tufillaro et al., 1992). As the task requires simple periodic bouncing, only periodic behavior was considered. Local linear stability analysis of this discrete map identifies a fixed-point attractor, when racket acceleration at impact $a_r$ satisfies the inequality (Schaal et al., 1996; Sternad, Duarte, Katsumata, & Schaal, 2000):

$$-2g\frac{(1 + \alpha^2)}{(1 + \alpha)^2} < a_r < 0$$

In this dynamically stable state, small perturbations of the racket or ball die out without requiring any corrections. To make this property of the task more intuitive, consider the illustration in Figure 2.5D: while hitting the ball with negative racket acceleration (the upwardly decelerating phase), a ball that contacts the racket earlier is hit with relatively higher velocity, which leads to a higher ball amplitude following the contact. Conversely, a ball that contacts the racket later, such as after an amplitude
overshoot, is hit with relatively lower velocity, which leads to a lower maximum ball height on the next bounce. Note that this purely physical relation automatically compensates for an over- or undershoot in the previous ball amplitude. This automatic compensation via covariation amongst the execution variables is the underpinning to the error compensation in dynamically stable performance.

This property leads to an interesting hypothesis for human control: If subjects establish such dynamically stable regime, they need not correct for small perturbations that may arise from the persistent motor noise. Thus, we hypothesized that subjects learn this “smart” solution of hitting the ball with negative racket acceleration to exploit dynamic stability. *This dynamically stable solution makes intrinsic noise matter less.*

### 2.6.4 Virtual Environment

In the experiment, the participant stood in front of a rear projection screen and was instructed to rhythmically bounce the virtual ball (white) to a target line (yellow) using a real table tennis racket (Figure 2.6A). The projected racket movements (red) were shown on the screen impacting the ball. One trial usually lasted 40 seconds. A light rigid rod with two hinge joints was attached to the racket and ran through a wheel, whose rotations were registered by an optical encoder. While the joints allowed the racket to move and tilt with minimal friction in all three dimensions, the encoder only measured the vertical displacement of the racket, in analogy with the model. Racket velocity was continuously estimated. The displacements of the virtual racket were controlled by the measured position of the real racket, and the vertical position of the virtual ball between impacts was determined using the ballistic flight equation. To simulate the haptic sensation of a real ball hitting the racket, a mechanical brake, attached to the rod, was activated at the ball-racket impact of each bounce. Racket acceleration at impact was analyzed after the
experiment and served as the primary measure of dynamic stability to test the hypothesis (K. Wei et al., 2007).

### 2.6.5 Experimental Findings

This hypothesis was tested in an experiment that involved 48 novice subjects, who performed 20 trials of 40 sec each (which corresponds to approx. 60 bounces per trial) (Ehrlenspiel, Wei, & Sternad, 2010). Figure 2.6B shows the medians over subjects in each trial. The error (red) declined with practice as to be expected. More importantly, this improvement was paralleled by a decrease in racket acceleration at impact (blue), starting from positive values at the beginning to assuming negative values after about 8 trials. This result highlights that the intuitive solution for novices is to accelerate to impart energy to the ball in the upward direction. Yet, this is not the most effective approach, when the ball is hit repeatedly with periodic racket movement. Further, from a biomechanical perspective hitting the ball with maximum velocity, i.e. at the moment of zero acceleration, would be the most energy-efficient solution. Maximum velocity would result in the highest ball amplitude for a given rhythmic racket trajectory. Nevertheless, subjects reliably converge to solution where the racket is in its decelerating phase at ball contact. Additional support for this strategy was revealed by estimating covariation amongst the three execution variables at impact (Ehrlenspiel et al., 2010). These results highlight that adopting solutions that afford dynamic stability is not trivial but is a smart solution that requires practice.

A series of studies tested the generality of this result in different experimental renderings to rule out the possibility that this could be the fortuitous effect of the tight constraints in the virtual set-up. The experimental set-ups ranged from having the participant manipulate a pantograph linkage that allowed precise control of the haptic
contact (Schaal et al., 1996; Sternad, Duarte, Katsumata, & Schaal, 2001), a real tennis racket to bounce a real ball attached to a weighted boom that confines the ball to a vertical path (Katsumata, Zatsiorsky, & Sternad, 2003; Sternad et al., 2001; Sternad, Duarte, et al., 2000), to having the participant freely bounce a real ball with a real tennis racket in 3D (Sternad et al., 2001; Sternad, Duarte, et al., 2000). The findings were robust: experienced performers hit the ball with negative racket acceleration. Further support that humans seek dynamically stable solutions came from a study by (Morice et al., 2007), where visuo-motor phase between the real racket and ball movements were shifted, performers still learned to exploit dynamic stability, where intrinsic noise matters less.

2.6.6 Interim Conclusions

Just as in the skittles task, participants were shown to find stable, or error-tolerant solutions in the redundant space of solutions to achieve the task. Further analysis of Covariation and Noise is still to come. The commonality of the two paradigms is that the subject’s control of the ball was confined to a very short moment. A first extension from this momentary interaction was made in an experiment that extended the duration of contact: a bean bag replaced the elastic ball, such that the contact duration became long enough to afford corrections and essentially became a catch followed by a resetting of initial conditions to throw (Katsumata et al., 2003). However, real-life skills include numerous skills that involve continuous interaction with objects. Hence, the next step is to design an experimental paradigm that features continuous interaction with a complex object: a cup of coffee.
2.7 Paradigm 3: Complex Object Manipulation, Carrying a Cup of Coffee

2.7.1 The Skill

Grasping a cup of coffee and leading it to one’s mouth is a seemingly mundane task. Yet, it exemplifies a class of movements that require continuous interaction with an object that has internal dynamics (Mayer & Krechetnikov, 2012). While one moves the cup, the coffee is only indirectly controlled via moving its container (Figure 2.7A). How the CNS controls interactions with such objects is intriguing, as the moving liquid creates time-varying reaction forces that have to be accounted for to keep the cup under control and not spill the coffee. The essential challenge of this task is to perceive these time-varying dynamics and establish a safety margin for transporting the cup of coffee.

2.7.2 The Model

Leading a cup of coffee to one’s mouth is a tall order to model mathematically. Hence, we again simplified the situation by modeling of the cup as a bowl with a ball moving inside it—representing the internal degree of freedom that affect the system’s dynamics (Hasson, Hogan, & Sternad, 2012; Hasson, Shen, & Sternad, 2012; Hasson & Sternad, 2014) as shown in Figure 2.7A. Assuming that the cup’s movements are confined to the horizontal direction and the ball can swerve within the cup, this reduced system is approximated by the well-studied cart-and-pendulum system (Figure 2.7B). The cart corresponds to the cup and the bob’s pendular movements represent the sloshing liquid. The governing equations of the system dynamics are:

\[(m + M)\ddot{x} = F\]

\[\ddot{\theta} = (\ddot{x}/l) \cos \theta - (g/l) \sin \theta\]
where $\theta$, $\dot{\theta}$, and $\ddot{\theta}$ are angular position, velocity, and acceleration of the ball; $x$, $\dot{x}$, and $\ddot{x}$ are the cart/cup position, velocity, and acceleration, respectively; $F$ is the force applied to the cup by the subject. Parameters of the system are $m$ and $M$ representing the masses of ball and cup, respectively; $l$ is the length of the rod (pendulum length); and $g$ is the gravitational acceleration. Hence, the model has four state variables $x$, $\dot{x}$, $\theta$, and $\dot{\theta}$ and the externally applied force $F$ that determines the behavior of the ball and cup system. The task is determined by five variables. The new problem is that only one variable (force) is under direct control of the subject, and even this variable is continuously influenced by the reaction forces of the ball. In fact, $F$ is the net force, as the ball or pendulum bob moves under the influence of the cart dynamics and gravity and imparts a time-dependent force to the cup in addition to the active force supplied by the actor. This makes the distinction into execution and result variables significantly more complicated than in the previous two examples.

2.7.3 Hypotheses

Nevertheless, the core feature of redundancy is present also in this task: Transporting the cup from one position to another can be achieved by multiple trajectories. How can one conceive of a result space determined by execution variables? One important constraint of the task is to keep the ball within the cup, which reduces the range of all available trajectories by limiting the forces that the actor can successfully apply. This task constraint can be expressed via the energy of the ball and its distance from escape (see more detail in Hasson, Shen, et al. 2012). Briefly, the current energy of the ball is a sum of its potential, kinetic, and inertial energy (energy conferred to the ball due to the acceleration of the cup). Potential energy is defined by the height of the ball
within the cup, kinetic energy is defined by the velocity of the ball, and the inertial energy is provided by the instantaneous acceleration of the cup (Hasson, Shen, et al., 2012). At each sampled point in the movement, one can define an energy margin as the difference between the current energy of the ball and the energy that would make the ball escape. This energy margin is the dynamic analogue to the safety margin, as defined in studies on grip force in object manipulation (Flanagan & Wing, 1997; Johansson & Westling, 1984). The energy margin is positive, when participants use forces that keep the current energy smaller than the energy level that would make the ball escape—a safe strategy. A more risky strategy would yield values of energy margin near zero, getting the ball close to the rim of the cup. Negative energy margin values indicate that the ball would escape in the near future, assuming no change in the currently applied acceleration.

When the task result is defined in these terms, execution space spans only three variables: acceleration of the cup/cart and position and velocity of the ball/pendulum. Figure 2.7C visualizes this execution space including a result surface that specifies the energy margin of zero. The area inside this manifold captures those executions that have positive energy margins (from 0 to 1), all solutions outside the manifold would lead to ball escape (assuming no change in acceleration). The plotted trajectory demonstrates an example evolution of the execution variables during a single trial of moving the cup. Note that the trajectory may leave the blue surface temporarily, but as long as participants quickly adjust their acceleration, they can prevent ball escape. However, larger deviations will lead to ball escape.
Extending previous arguments, that *performers acquired solutions with stability to make intrinsic noise matter less*, we tailored this hypothesis to the current task. As subjects moved the cup from A to B, and not “spill the coffee” (lose the ball), we hypothesized that with practice subjects will seek solutions that have large energy margins. Further, we hypothesized that individuals with greater movement variability, or intrinsic noise, would seek solutions that were less risky (and vice versa).

### 2.7.4 Virtual Environment

The cart-pendulum system was rendered in a virtual environment where participants only saw the bob of the pendulum and the arc of its trajectory that represented the cup (Figure 2.7D). Participants controlled the virtual cup using a robotic manipulandum that was constrained to move the cup in the horizontal direction (Haptic-Master, Moog, Netherlands). The task goal was to transport the virtual cup from an initial position to a target position at 40 cm distance (green boxes). In two conditions, subjects moved the cup either as fast as possible (minimum-time group) or at the specified time of 2 seconds (target-time group). The time window was signaled to subjects by a white box that began descending onto the target box at constant velocity upon leaving the start box. Two seconds passed when the white box was exactly overlaid on the target box.

### 2.7.5 Empirical Findings

Two groups of 9 participants each practiced this task for 300 trials in the two conditions (Hasson, Shen, et al., 2012). The motivation for the target-time condition was to create a condition where there was redundancy: two seconds gave subjects ample time to develop different strategies. We hypothesized that they would seek “*stable strategies*” that optimized the energy margins to *allow for noise and perturbations to matter less*. For comparison, the minimum-time condition only presented one optimal solution. To
increase speed, subjects were expected to decrease their energy margins. In the two conditions changes in both directions would signal that subjects were aware of their variability.

With practice all participants in the target-time group became better in placing the cup in the target box in the specified 2-second interval; similarly, the minimum-time group decreased their movement time. Further, in both conditions participants lost the ball less often and their trial-to-trial variability decreased. The most central variable was the energy margin, which was calculated for each point along the cup trajectory, yielding a time-varying pattern. Figures 2.8A-B exemplify the pattern of results in both conditions and their change with practice. In the target-time condition, participants increased their energy margin, indicating that strategies became safer with practice. This result is consistent with the hypothesis, because higher energy margin values provide more tolerance to intrinsic noise. In contrast, the minimum-time group decreased their energy margin, indicating that their strategies became more risky as a compromise to increase speed. Yet, the number of ball losses still decreased, reflecting that subjects were sufficiently sensitive to their variability. Figure 2.8C summarizes the change in energy margin with practice in all individuals in the two conditions.

To test the second hypothesis stating that the preferred energy margin should be adjusted to each individual’s variability, we examined the correlation between the variability of the total energy throughout the movement and the energy margin during that same movement in the last 30 trials of the experiment (for details see Hasson, Shen, et al., 2012). The results showed that participants in the minimum-time group (blue circles), who had more variability also showed greater energy margins, suggesting that
more variable individuals used safer movement strategies (Figure 2.9). However, this correlation was not present in the minimum-time group (red triangles). One explanation is that the variability in the total energy did not simply reflect the variability at the ‘physiological limit’, but also captured the greater variability in the movements when less constrained. Despite this partially inconsistent result, the findings lend support to the hypothesis that actors take stability properties of the task and their own intrinsic noise characteristics into account when learning to manipulate a “cup of coffee”.

2.8 Conclusions and Outlook

This chapter aimed to present a methodological approach to study the acquisition and control of novel and more complex skills that capture coordination challenges inherent to real-life motor skills. This work differs from much of the more recent research on motor learning or adaptation, where an existent movement pattern is adapted to external manipulations. We advocated to study more complex tasks to understand the acquisition of novel skills, which is an essential and ubiquitous process that enables humans not only to perform extraordinary actions, such as skateboarding, but also to eat with knife and fork or drive a car. How the brain changes and stores such infinitely many new skills that essentially define our human existence remains one of the core open questions in human neuroscience.

We proposed that the minimal ingredient for skill is redundancy, in particular extrinsic or task redundancy that allows individuals to make choices or even decisions about how they perform the task. This contrasts with much of the previous work in motor neuroscience that has tended to focus on simple and well-defined movements with a single goal. These experimental paradigms are well suited to study simpler forms of
learning, including habituation and sensitization, classical and operant conditioning and 
motor adaptation. In order to extend the realm of scientific study and address more high-
level skills, we developed a methodological approach that still afforded hypothesis-driven 
research. In three examples, we exemplified our model-based approach.

All three experimental paradigms involved interaction with an object and thereby 
exemplified different aspects of tool use, a fundamental human skill. They form a logical 
progression in the physics they involve: In skittles, the trajectory of the thrown ball is 
fully determined at the moment of release; there is no further interaction with the ball. 
Every trial is a new start, setting new initial conditions. In ball bouncing, subjects again 
contact the ball at a very brief moment that fully determines the subsequent ball 
trajectory. However, unlike in skittles, the successive ball contacts are not independent: 
the ball’s trajectory from the previous impact influences the next ball impact. Lastly, in 
the cup-and-coffee task, the force applied to the cup and the force exerted by the sloshing 
coffee onto the cup are in continuous and complex interplay. In this continuous 
dynamical system, every moment in human control matters. Due to this increasing 
complexity, the methods for analysis in the three paradigms need to be different. 
Nevertheless, even though the models are different and render different variables, all 
three experiments pursued one core hypothesis: Humans seek stable solutions where 
intrinsic noise matters less.

Each experimental task was designed to allow an infinite number of solutions to a 
given task. Given these multiple options, it is necessary to have a model as reference to 
evaluate the observed variability within subjects across practice, but also across subjects. 
It is essential that the model described not only the physics of the task, but also how the
subject interacted with the physics, i.e., how the subject executed the task to achieve a result. If the functional relationship between execution and result is known, the space of all solutions is known and the manifold of zero error solutions can be calculated. With this knowledge not only can the practice-induced progression in performance be characterized, but individual differences can also be studied in a more principled way. For example, a person with high variability may prefer safer strategies, while a person with low intrinsic noise may prefer a more risky strategy. While this is a first start, more work is needed to better understand inter-individual differences (Cesqui, d'Avella, Portone, & Lacquaniti, 2012).

While our current research has focused on developing and testing the approach on healthy individuals, the experimental platforms are ideal to take the next step: diagnose coordination deficits and design interventions to accelerate learning and modify behavior of patient populations. For example, the virtual rendering of the task lends itself to perceptual modifications that may shape the characteristics of behavior. One study exemplifies the initial step in this direction: using a modified skittles task, we showed how a decrease in perceived variability could affect movement strategies in children with dystonia (Chu, Sternad, & Sanger, 2013). Following such first forays, the arena of questions and applications is wide open.
2.9 Figures

Figure 2.1. (A) Motor adaptation paradigm. The participant starts with a highly learned movement, such as reaching, which is performed with zero error in the pre stage (red). A perturbation is applied during a block of trials (yellow), until the participant adapts and re-establishes the initial performance. When the perturbation is removed, there is a short after-effect, indicating that something has been learned. However, the original movements is resurrected quickly (blue); thus there is no net change in skill. (B) Skill acquisition. In the pre-stage, the participant performs a novel motor skill with high error (red). The participant decreases error over practice (yellow). Error in the retention stage (blue) is less than the error in the pre stage (red), reflecting the acquisition of the novel skill.
Figure 2.2. (A) Real-life skittles game. (B) Two-dimensional mechanical model of skittles. The ball (white circle) is attached to two orthogonal springs centered at the origin of the center post (red circle). The ball trajectory (dotted line) upon release is determined by the springs and depends on two execution variables: release angle \( \phi_r \) and velocity \( v_r \) of the ball. Error is defined as the minimum distance between the ball trajectory and the target center (yellow circle). (C) Execution space for four different target locations. Error is calculated for each point in execution space and indicated in color; the lighter shades denote lower errors. The black line represents the one-dimensional solution manifold. (D) Execution space with non-linear solution manifold with two data sets representing 20 throws collected on the first day (blue) and after 11 days of practice (yellow).
Figure 2.3. (A) Participant interacting with the virtual environment in the skittles setup. The center post (red), and target (yellow) and the movements of the lever arm (magenta) and the ball (white) are presented to subjects on the rear projection screen in front of them. (B) Mean T- (red), N-Cost (green), and C-Cost (blue) for three expert participants over 45 blocks (15 days) of practice. Error bars show standard error across participants.
Figure 2.4. (A) Execution space with two sets of arm trajectories representing 20 trials collected on the first day (blue) and after 11 days of practice (yellow). (B) Mean timing error (green) and time in hit zone for three expert participants over 15 days of practice. Error bars show standard errors across participants.
Figure 2.5. (A) Real-life rhythmic ball bouncing skill. (B) Model of the racket-ball system. The vertical ball position between each instantaneous impact follows ballistic flight, which depends on three execution variables: ball $v_b^-$ and racket $v_r^-$ velocities just before impact, and racket position $x_r$ at impact. Error is defined as the absolute difference between the target height and max ball height. (C) Execution space of the racket-ball system. The gray plane represents the two-dimensional solution manifold. Two data sets represent bounces in first trial (blue) and after 20 trials of practice (yellow). (D) Illustration of dynamic stability associated with negative acceleration. The black trajectory of the ball exemplifies stable performance. A positive perturbation to racket velocity just before impact was added to the green ball trajectory at the arrow. The error from the perturbation is corrected after three bounces without any changes in the sinusoidal racket trajectory.
Figure 2.6. (A) Participant interacting with the virtual environment. The movements of the racket (red) and ball (white), and the target height (yellow) is presented to subjects on the large screen in front of them. (B) Median error (red) and racket acceleration at impact (blue) for 48 novice participants over 20 trials of practice. Error bars show the 25th and 75th percentiles across participants.
Figure 2.7. (A) Holding a cup of coffee simplified as holding a ball in cup. (B) Model of the cart-and-pendulum system. In the mathematical model, \( x \) is the cart/cup position, \( \theta \) is the ball angular position, \( M \) is the cup mass, \( m \) is the ball mass, and \( l \) is the pendulum length. (C) An exemplary successful trajectory (trial) in 3-dimensional execution space, spanned by ball angle and velocity and the acceleration of the cup, which together specify the energy margin of the ball. The energy margin describes a 2-dimensional manifold (meshed blue surface). Assuming the cup acceleration is unchanged, the trajectory inside the manifold has a positive safety margin and the ball will remain in the cup (green trajectory); if the trajectory is outside (black trajectory), the energy margin is negative and the ball will escape unless a correction is made. (D) Participant interacting with the virtual environment. A haptic manipulandum provides the mechanical interaction with the object model and the movements of ball and cup appear on the rear-projection screen. The task is to transport the cup from the left green start box to the right green target box as fast as possible (minimum-time condition) or in exactly 2 s (target-time condition) without letting the ball escape from the cup.
Figure 2.8. Energy margin over the time of a trial (normalized to 100%) for an exemplary participant from the (A) target-time group and the (B) minimum-time group. Dark/light shading indicates a decrease/increase in energy margin between the average for first vs. last 30 trials. (C) Summary changes in energy margin with practice. Each data point represents the average for the first (early practice) or last (late practice) 30 trials for the target-time (triangles) and minimum-time group (circles).
Figure 2.9. Correlations between energy margin and trial-to-trial variability (standard deviation) of total energy during the last 30 trials for the target-time group (triangles) and the minimum-time group (circles).
3. Skilled Motor Performance: Overcoming Limits in Temporal Precision from Neural Noise

In Chapter 2, task redundancy (i.e. multiple solutions that lead to task success) was proposed as the essential ingredient for an experimental task to allow for the study of complex skill. In tasks with only a single optimal solution, reducing error and variability is the only strategy available to improve task performance. When there is more than one possible solution, however, learning strategies can be more complex. This study examines the learning strategies used for two versions of a virtual throwing task with redundancy. Specifically, it examines how the structure of task redundancy affects learning strategies during long-term practice. Particular attention is given to timing, an element regarded critical for accurate throwing. These experiments will show that timing is not always as critical as commonly believed. Rather, a redundant task can also affords shaping the arm trajectory to create windows that are tolerant to timing variability.

I contributed to the experimental design, data collection, data analysis, statistical testing, and interpretation of results.

3.1 Summary

Accurate throwing is one exemplary motor skills that reportedly requires extremely precise timing of the ball release (Calvin, 1983; Chowdhary & Challis, 1999). Accuracy and precision of timing in motor behavior is of long-standing interest as the limitations shed light on the information transmission in the nervous system. Naturally, reducing variability in timing is presumed key for improving accuracy and precision (Hore, Watts, & Tweed, 1996; Timmann, Watts, & Hore, 1999). However research has also shown that
motor tasks can have redundancy (i.e. multiple solutions that achieve task success), offering strategies that are tolerant to variability in behavior (Cohen & Sternad, 2009, 2012). The aim of this study was to determine whether humans improve throwing performance by reducing timing variability or rather increasing timing tolerance to overcome limitations of nervous system.

Subjects practiced a virtual throwing task for 11 daily sessions. While subjects learned to both decrease timing variability and increase timing tolerance, there was no correlation between timing variability and performance after the first practice day. Rather, those who were more tolerant to timing errors had better task performance. To further scrutinize this result, the task was modified such that the structure of the task redundancy obviated the ability to increase timing tolerance. Consequently, the only means of achieving a better task result was to lower timing variability.

These results demonstrate how different learning strategies are used based on the solution space of task. Some of the exquisite coordinative skill may not achieved by reducing variability alone, but by also circumventing physiological limitations and adopting strategies that make intrinsic variability matter less.

3.2 Results

Skilled motor performance requires many different types of timing (Zelaznik, Spencer, & Ivry, 2008) – e.g., timing the interval between consecutive actions (Wing & Kristofferson, 1973a, 1973b) and timing an action in response to a stimuli (Poulton, 1957). A less studied form of timing is timing an event during the execution a complex movement, like timing the release of a ball or dart during a throw. To achieve successful performance in a motor skill like throwing, precise timing of highly coordinated
movements appears to be critical. In fact, an argument has been made that learning to accurately throw projectiles has provided humans with an evolutionary advantage (Calvin, 1983).

Prior studies have reported that in throwing, the variability in timing the release of the projectile is approximately 9-12 ms (Cohen & Sternad, 2012; Müller & Loosch, 1999). However, levels of timing variability as low as 1-2 ms have also been reported (Calvin, 1983; Chowdhary & Challis, 1999; Hore & Watts, 2011; Nasu, Matsuo, & Kadota, 2014; Smeets et al., 2002). It is unclear whether the central nervous system can achieve such temporal acuity as these numbers exceed timing errors and variability reported in more controlled laboratory tasks, such as tapping where timing variability is reportedly ~10ms (R. M. C. Spencer, Zelaznik, Diedrichsen, & Ivry, 2003; Sternad, Dean, & Newell, 2000). Previous studies have also shown that the variability in release timing is positively correlated with variability in throwing performance (Hore et al., 1996). While this finding further suggests that reducing variability in timing is key for improving throwing performance, evidence indicates that humans also learn strategies that may increase their tolerance to timing variability for improving task performance (Cohen & Sternad, 2012; Müller & Loosch, 1999). Given that humans are only able to reduce their intrinsic neuromotor noise to a limited degree (Faisal et al., 2008), they may shape their behavior such that variability in execution matters less (Cohen & Sternad, 2009; Müller & Sternad, 2004b; Sternad et al., 2014).

This study aimed to determine the strategies humans employ to compensate for their irreducible noise and nevertheless achieve high “temporal” accuracy. Common experimental paradigms used to study other types of timing, such as interval, duration,
and response timing, are highly controlled to reduce the task to timing alone. More complex tasks, however, have a space of solutions that afford different options to satisfy tight timing demands. Addressing this aim sheds light on how humans acquire and perform complex motor skills that outwardly appear to have high demands for precise timing.

Specifically, we hypothesized that in throwing, subjects who develop trajectories that are more tolerant to timing errors will have better task performance, provided that the task affords developing such timing-tolerant trajectories (Hypothesis 1A). Hence, there would be no relation between timing variability and task performance (Hypothesis 1B), which would be contrary to previous results (Hore et al., 1996). To test these hypotheses, we examined 12 subjects as they learned a virtual throwing task. Each subject practiced the task for 11 days, with 240 throws per day, yielding a total of 2640 throws collected over 2-3 weeks. This extended practice aimed to assess learning strategies until skilled performance reached an asymptote, which is unique compared to other studies of throwing that examine performance within a single practice session (Nasu et al., 2014; Smeets et al., 2002).

In the virtual throwing task, the subject throws a tethered ball around a post to hit a target skittle at the far side (Figure 3.1A). Subjects manipulated a horizontal lever arm with a single-joint rotation about the elbow to throw a virtual ball in a 2D virtual workspace (Figure 3.1B-C). They were instructed to throw the virtual ball such that it travelled through the center of the virtual target without hitting the center post. At the start of each throw, the subject grasped a wooden ball affixed to the distal end of the lever arm and pressed a force sensor with his or her index finger. This attached the virtual ball
to the tip of the virtual lever arm. The subject then rotated the lever arm with his/her forearm and triggered the ball release by lifting the index finger off the force sensor. Upon release, subjects saw the ball traverse an elliptical path in the virtual workspace for 1.5s (Figure 3.1C). Error was defined as the minimum distance between the ball path and the target center. If the magnitude of error was below the success threshold of 1.1cm, the target changed color from yellow to green to indicate a successful hit.

The trajectory of the virtual ball, and consequently error, was determined by two execution variables: the angular position and velocity of the lever arm at the time of release. Figure 3.1D illustrates the solution space of the task, which is the mapping between these two execution variables onto the result variable, error, for the target location shown in Figure 3.1C. For each potential combination of angular position and velocity at release, an error value was calculated from the motion equations for the ball path and then marked by color (darker shades indicate increasing error). Errors outside the target boundary (>2.5cm) are shown in black, and errors inside the target radius, but greater than the success threshold are shown in yellow (2.5cm > error > 1.1cm). The area shown in green is referred to as the “success manifold” and denotes errors inside the success threshold that result in a target hit (<1.1cm). Note that this task is redundant as there are infinite pairings of angular position and velocity that lead to successful target hits (Cohen & Sternad, 2012; Dupuy, Mottet, & Ripoll, 2000; Kudo, Tsutsui, Ishikura, Ito, & Yamamoto, 2000).

Figure 3.1D shows two arm trajectories from one example subject plotted as sequences of angular position and velocity pairings. The red trajectory from Day 1 exemplifies a trajectory that intersects the success manifold, and precise timing of the
release at the intersection is needed to successfully hit the target. Conversely, the trajectory shown in blue from Day 11 travels within the success manifold for an extended period of time. This means that multiple release moments on the arm trajectory would achieve a successful hit. Thus, variability in the timing of ball release has little influence of task performance and makes this trajectory tolerant to timing errors.

3.2.1 Influence of Timing Tolerance on Error Variability

To assess the influence of timing tolerance on task performance, we first examined how these behavioral features changed over 11 days of practice. Two measures, task success and error variability, were used to characterize overall task performance. Task success was defined as the percentage of throws with error below the success threshold on each practice day. Error variability was defined as the interquartile range (IQR) of error on each day. A one-way repeated-measures ANOVA with practice day as the within-subjects factor was conducted on each of the measures. As to be expected, task success significantly increased with practice ($F(4.03, 40.27)=27.71, p<.001$; Figure 3.2A) from Day 1 ($M=10.0\%, SD=3.97\%$) to Day 11 ($M=24.48\%, SD=6.39\%$), and error variability significantly decreased ($F(1.49, 16.34)=18.22, p<.001$; Figure 3.2A) from Day 1 ($M=12.70cm, SD=6.27cm$) to Day 11 ($M=4.64cm, SD=1.16cm$).

Timing tolerance was defined as the median timing window on each day of practice characterized. The timing window for each throw was quantified as the amount of time a given arm trajectory travelled within the success manifold (Figure 3.1E). Subjects also significantly increased their timing tolerance over practice ($F(3.58, 39.32)=13.05, p<.001$; Figure 3.2B) from Day 1 ($M=9.00ms, SD=3.59ms$) to Day 11 ($M=16.25ms, SD=5.17ms$). Figure 3.2C illustrates how an example subject’s arm trajectories became more tolerant from the first practice day to the last.
To test the hypothesis that timing tolerance is associated with lower error variability, Pearson linear correlations between the two measures were computed for each subject on each practice day (Table 3.1A). One Day 1, no significant relation between timing tolerance and error variability was observed ($r=-.24, p=.46$; Figure 3.2D). However from Day 2 onwards, a significant negative correlation between these two measures ($r_s>.58, p_s<.048$) signaled that subjects with longer timing windows also had lower variability in error (Figure 3.2E). This latter finding provided strong support for Hypothesis 1A, with the important caveat that this relation did not emerge until the second day of practice.

### 3.2.2 Influence of Timing Variability on Error Variability

Given that subjects learned strategies tolerant to timing errors, we predicted that variability in timing error would not influence task variability (Hypothesis 1B). Timing error of each throw was defined as the difference between the time of actual release and the ideal release time in a given arm trajectory (Figure 3.1F). The ideal release time was defined as the time where a given release would have yielded the smallest error. *Timing variability* was characterized by the IQR of timing errors on each practice day.

A one-way ANOVA revealed that timing variability significantly decreased over practice ($F(4.24, 46.63)=11.38, p<.001$; Figure 3.2B) from Day 1 ($M=35.98\text{ms}, SD=6.65\text{ms}$) to Day 11 ($M=25.52\text{ms}, SD=4.22\text{ms}$). Notwithstanding this decrease, subjects did not reduce their timing variability to a level reported in prior studies (Calvin, 1983; Chowdhary & Challis, 1999; Hore & Watts, 2011; Nasu et al., 2014; Smeets et al., 2002). Even after 11 practice days, subjects only reduced their timing error variability to a group average of 25.2ms (Figure 3.2B). Pearson linear correlations between the timing variability and error variability were computed for each practice day (Table 3.1B). On
Day 1, a significant positive correlation signaled that subjects with higher timing variability tended to have higher error variability \((r=0.87, p<.001)\) (Figure 3.2C). This finding was not surprising given the lack of correlation between timing tolerance and error variability on Day 1. However, the significant correlation between timing tolerance and task variability from Day 2 onwards was no longer coupled with a significant correlation between timing and error variability \((p>.15, \text{Figure } 3.2D)\). This latter finding was consistent with Hypothesis 1B.

The correlation analyses of timing variability and tolerance with error variability revealed that after the first practice day, the tolerance to timing errors in the arm trajectories was the critical determinant in this task, as hypothesized (Hypotheses 1A and 1B). Though it is important to recognize that the results of Day 1 did not initially support this prediction. Rather, they supported the opposing claim that the variability of timing errors is related to task performance. However, this result only emphasizes that subjects likely had to adopt the strategy of using a more timing-tolerant arm trajectory to increase performance over extended practice.

### 3.2.3 Task-Dependent Nature of the Influence of Timing Variability and Tolerance

Not all tasks, however, may afford this tolerance to timing. This could be due to the lack of task redundancy (i.e. only one solution achieves task success), or due to biomechanical or other physical constraints that do not allow subjects to exploit the redundancy of task. One would expect that the absence of tolerant strategies heightens the need to decrease timing variability for improving performance.

An example of such a task is shown in Figure 3.3A. This task is the same virtual throwing task, however the target was moved to a different location, which changed the topography of the result space (Sternad et al., 2014). While the previous constellation
allowed subjects to align their arm trajectory with the success manifold (Figure 3.1D), the new target location yielded a vertical success manifold, meaning that task success could only be achieved with a limited range of angular positions (Figure 3.3B). As it is not possible for subjects to maintain the same angle with non-zero velocity, it is impossible for subjects to align their arm trajectory with the success manifold.

We hypothesized that in tasks where subjects could not develop tolerant trajectories, those who achieved lower timing variability will have better task performance (Hypothesis 2). To test this hypothesis, we assessed subjects that practiced for three days on the target configuration shown in Figures 5.3A-B. Over practice, subjects significantly improved task success ($F(1.82, 18.24)=5.75, p=.013$; Figure 3.3C), achieving high task performance by Day 3 ($M=81.14\%, SD=13.39\%$). Subjects also significantly decreased error variability ($F(1.51, 15.10) = 23.43, p<.001$; Figure 3.3C) from Day 1 ($M=2.97cm, SD=1.10cm$) to Day 3 ($M=1.79cm, SD=0.72cm$) as well as timing variability ($F(1.32,13.16)=12.95, p=.002$, Figure 3.3D) from Day 1 ($M=52.23ms, SD=11.00ms$) to Day 3 ($M=35.91ms, SD=16.29ms$). As expected, subjects did not significantly increase their timing window ($F(1.38,13.77)=.083, p=.85$, Figure 3.3D) from Day 1 ($M=47.05ms, SD=4.52ms$) to Day 3 ($M=47.82ms, SD=9.03ms$), reflecting that subjects were in fact unable to increase timing tolerance in this task.

Pearson correlations between task error variability and timing error variability on each practice day did not detect any significant relationships between timing tolerance and error variability (Day 1: $r=-.44, p=.18$; Day 2: $r=.52, p=.10$; Day 3: $r=.33, p=.32$; Figure 3.3E). As hypothesized, there were significant positive correlations between
timing and error variability on all days of practice (Day 1: $r=.79$, $p=.0035$; Day 2: $r=.97$, $p<.001$; Day 3: $r=.89$, $p<.001$; Figure 3.3E).

3.3 Discussion

There are many sources of noise in the nervous system, from the subcellular level to organ level of muscles (Faisal et al., 2008). Even temporal precision of neuron spike timing is only accurate to the order of milliseconds (Ikegaya et al., 2004; Mainen & Sejnowski, 1995), which explains timing variability in behavior has been observed to be at least ~9ms (Cohen & Sternad, 2012; R. M. C. Spencer et al., 2003; Sternad, Dean, & Newell, 2000). To overcome these limitations in timing precision, the results of these two experiments suggest that when the task provides the opportunity to develop a timing-tolerant arm trajectory, the extent to which subjects exploit this strategy determines their performance. When the task does not afford this strategy, however, the ability to reduce timing variability becomes the determining factor of task performance. These results offer a possible explanation for contradictory results on release timing during throwing (Cohen & Sternad, 2012; Hore et al., 1996; Smeets et al., 2002): just because a task has redundancy, does not necessarily mean that learners can take advantage of it. More importantly, the results have several important implications that extend beyond learning to improve throwing performance.

The results of this study also emphasize the importance of a thorough understanding of the experimental task. Previous studies have used sensitivity analyses to assess tolerance to timing variability, however these local analyses alone can only inform the degree to which an individual is using a trajectory that is tolerant to error and intrinsic noise (Calvin, 1983; Chowdhary & Challis, 1999; Smeets et al., 2002). They do not
inform to what extent the topography of the task’s redundancy affords a timing tolerant trajectory given the biomechanical and physical constraints of the learner. By using a virtual throwing task, this study that permitted a prior analysis of the solution space to make predictions about the timing and noise tolerant performance.

Importantly, this analysis not only revealed that the influence of timing variability is task-dependent, it also revealed why: the topology of the success manifold. As demonstrated in this study, slight changes in the task configuration, such as target placement, can substantially affect the strategies learners can use to improve performance. Given that most studies on throwing use slightly different experimental paradigms, it is not surprising that they arrive at different conclusions on the importance of timing (Cohen & Sternad, 2012; Jegede, Watts, Stitt, & Hore, 2005; Nasu et al., 2014; Smeets et al., 2002). While some studies have used task models to assess learning strategies (Nasu et al., 2014; Smeets et al., 2002), these models require approximations about the projectile during flight. Given the many assumptions, these task models did not permit a full analysis as is possible for the fully controlled virtual task.

This level of task analysis is desirable not only for throwing, but for all movements as they are not performed in a void but situated in the external workspace. The present findings show that learners are sensitive to the redundancy of a task and utilize strategies to improve performance based on the topography of the task redundancy. While reducing variability is undoubtedly important for learning, concurrent efforts to increase tolerance to variability can be just as critical, if not more so, for performing highly complex skills. And yet, while the relation between neural and behavioral variability is actively investigated (Churchland, Yu, Ryu, Santhanam, &
Shenoy, 2006; Churchland, 2015; Renart & Machens, 2014), there is a very limited understanding of the mechanisms through which the CNS learns to identify and exploit the redundancy of tasks commonly found in everyday life. The use of complex, yet controlled virtual tasks, along with neurophysiological recordings, provides the opportunity to enhance this understanding in the future research.

3.4 Methods

3.4.1 Subjects

3.4.1.1 Experiment 1

12 healthy, right-handed students (5 females and 7 males, mean age 20.8±3.2 years) from Northeastern University took part in this experiment. Each subject practiced the virtual throwing task (target location: 60cm above and 60cm to the left of the origin; Figure 1B) for 11 days, with 240 throws per day, yielding a total of 2640 throws collected over 2-3 weeks.

3.4.1.2 Experiment 2

11 healthy, right-handed students (7 females and 4 males, mean age 20.9±2.4 years) from Northeastern University took part in this experiment. Each subject practiced the virtual throwing task (105.5cm above and 5cm to the right of the origin; Figure 3A) for 3 days, with 240 throws per day, yielding a total of 720 throws collected over 1 week.

None of the subjects had any prior experience with the experimental task. All subjects gave informed written consent before the experiment and received monetary compensation upon completion. The experimental protocol was executed in compliance with the Institutional Review Board of Northeastern University.
3.4.2 Task and Apparatus

The experimental task was modeled after a British pub game named skittles. In skittles, the player throws a tethered ball around a center post to hit a target skittle on the other side of the post (Figure 3.1A). In the experimental task, the subject manipulated a horizontal lever arm with a single-joint rotation about the elbow to throw a virtual ball in a 2D virtual environment (Figure 3.1B). The virtual environment consisted of a top down view of the skittles work place and was projected onto a large screen .6m in front of the subject (Figure 15.C). On the screen, subjects saw the post (radius = 25cm) centered at the origin, the ball (5cm), and a virtual lever arm (length = 40cm) with the axle located 150cm below the origin. In Experiment 1, the target (radius = 5cm) was located 60cm above and 60cm to the left of the origin. In Experiment 2, the target was located 105.5cm above and 5cm to the right of the origin. The positions and size of each object was defined in a virtual coordinate system, and then scaled up to project the graphics on the large screen.

Subjects were instructed to throw the ball such that it travelled through the center of the target without hitting the post. At the start of each throw, the subject grasped a wooden ball affixed to the distal end of the lever arm and pressed a force sensor (Interlink Electronics, Camarillo, CA) with his or her index finger. This attached the virtual ball to the end of the virtual lever arm. The subject then swung his or her arm and triggered the ball release by removing the index finger off the force sensor. Upon release, all subjects saw the ball traverse an elliptical path in the virtual scene for 1.5s. The path of the ball was fully determined by the angular position and velocity of the lever arm at the moment of release. Error for a given throw was defined as the minimum distance between the ball path and the target center. If the ball path travelled between the origin and the target
center, error was considered negative; otherwise the error was considered positive. If the magnitude of error was below 1.1cm, the target changed color from yellow to green to signal a successful target hit.

### 3.4.3 Task Model

A two-dimensional model in which the ball was attached to the origin by two orthogonal massless springs was used to generate the ball path from each throw. The equations for the ball position in the x- and y- directions at time $t$ are:

$$x(t) = A_x \sin(\omega t + \varphi_x) e^{-\frac{t}{\tau}}$$

$$y(t) = A_y \sin(\omega t + \varphi_y) e^{-\frac{t}{\tau}}$$

The frequency $\omega$ denotes the natural frequency of the system, and relaxation time $\tau$ was used to introduce dampening to approximate realistic behavior. The amplitudes $A_x$, $A_y$, and phases $\varphi_x$ and $\varphi_y$ were calculated from the angular position and velocity of the ball at release based on the recorded movement of the lever arm. Due to the restoring forces proportional to the distance of the ball from the origin, the ball was accelerated toward the origin.

### 3.4.4 Measurement of Execution Variables

#### 3.4.4.1 Experiment 1

All subjects practiced the same experimental task, however the experiment was conducted in two rounds using different measurement devices. Five subjects (2 females and 3 males, mean age 19.2±.4 years) participated in the first round of data collection. The angular position of the lever arm was measured with an analog potentiometer (Vishay Spectrol, Ontario, Canada). Signals from the potentiometer and force sensor were sampled at variable rate of ~700Hz using a DT300 data acquisition board (Data
Translation, Inc., Marlboro, MA). The angular velocity of the lever arm was calculated online using linear regression over the 20 samples of time and angular position prior to release. The angular position and velocity measurements were spline-fitted and then resampled at 1kHz post-hoc in order to compute the dependent measures.

In the second round of data collection, seven subjects (3 females and 4 males, mean age 22.0±3.8 years) performed the same task with updated experimental equipment. The angular position of the lever arm was measured with a digital encoder (BEI Sensors, Goleta, CA), and the signals from the encoder and force sensor were sampled at consistent rate of 1kHz using an NI USB-6229 BNC data acquisition board (National Instruments, Austin, TX). As in the previous round, the angular velocity of the lever arm was calculated online using linear regression over the 20 samples of angular position prior to release. Because the angular position and velocity measurements were already acquired at 1kHz sampling rate, no post processing was required to compute the dependent measures.

A 2 (Experiment Round) x 11 (Practice Day) ANOVA with experiment round as a between-subjects factor and practice day as a within-subjects factor was conducted on each dependent measure to ensure that that difference in angular position sensors and sampling rate did not influence the calculation of the measures. The Greenhouse-Geisser correction factor was applied to the within-subject effects (Kirk, 1995). As mirrored in the results section, there was a significant effect for practice day on all dependent measures (ps < .001). However, there was no significant effect of experiment round on any of the dependent measures (task success: $F(1,10) = .024, p=.88$, task variability: $F(1,10) = 1.63, p=.23$, timing error variability: $F(1,10) = 3.02, p=.11$, timing error
insensitivity: $F(1,10) = .79, p=.50$), nor a significant interaction on any measure (task success: $F(10,100) = .61, p=.66$, task variability: $F(10,100) = .83, p=.44$, timing error variability: $F(10,100) = 2.52, p=.08$, timing error insensitivity: $F(1,10) = 1.53, p=.22$).

Because the data from the two experiment rounds was not statically different, the data from both rounds was combined for the remaining statistical analyses.

### 3.4.4.2 Experiment 2

All subjects performed the same task with updated experimental equipment as described above. The angular position of the lever arm was measured with a digital encoder (BEI Sensors, Goleta, CA), and the signals from the encoder and force sensor were sampled at consistent rate of 1kHz using an NI USB-6229 BNC data acquisition board (National Instruments, Austin, TX).

### 3.4.5 Dependent Measures

#### 3.4.5.1 Task Performance

Task success, defined as the percentage of throws with error below the success threshold, and error variability, defined as the interquartile range (IQR) of error, of each day characterized task performance.

#### 3.4.5.2 Timing Variability and Tolerance

For each throw, timing error and timing window were computed. To calculate these measures, the arm trajectory was first converted from a sequence of angle-velocity pairs to an error signal by calculating the error resulting from the angle-velocity pair at 1ms resolution. The timing error of each throw was defined as time of actual release was compared to the ideal release time for a given arm trajectory (Figure 3.1E). The ideal release time was defined as the time in the arm trajectory with the smallest error magnitude. In cases where there were two possible ideal release times as exemplified by
the red trajectory in Figure 1D, the ideal release time closest to the actual release time was used. Timing variability was characterized by the IQR of timing errors in each practice day.

Timing tolerance was quantified by the median timing window of trajectories in each day of practice. The timing window for each throw was defined as the amount of time error magnitude of the arm trajectory was less than success threshold (Figure 3.1F). An increase in timing window would indicate that the arm trajectory became more tolerant to timing errors.

Prior to any statistical analyses, all dependent measures were checked for deviations from normal distributions using W-tests (Shapiro & Wilk, 1965). Medians and interquartile ranges were used because the deviations in the measures within each practice day were not strictly Gaussian.

3.4.6 Statistical Analyses

To characterize how the dependent measures changed with practice, a one-way repeated measures ANOVA with practice day as a within-subjects factor was conducted on each dependent measure over the all days of practice for each experiment. The Greenhouse-Geisser correction factor was applied to the within-subject effects (Kirk, 1995). Pearson linear correlation coefficients between timing tolerance and error variability as well as timing variability and error variability were computed for each practice day.
3.5 Figures and Tables

Figure 3.1. Task description for the skill learning experiments. (A) Real skittles throwing task. The experimental task was modeled after the British pub game skittles. (B-C) Experimental setup of virtual throwing task. (D) Solution space of modeled skittles task with two example arm trajectories. White circles represent the points of release. (E) The timing window for each throw was quantified as the amount of time a given arm trajectory travelled within the success manifold. (F) Timing error of each throw was defined as the difference between the time of actual release and the ideal release time in a given arm trajectory.
Figure 3.2. Experiment 1 Results. (A) Error variability and task success across practice. (B) Timing variability and timing tolerance across practice. Shaded regions represent ± 1 standard error of the mean. (C) Example arm trajectories in solution space from Day 1 (red) and Day 11 (blue). White circles represent the points of release. (D) Correlations between timing variability and error variability (blue circles) and timing tolerance and error variability (red squares) on Day 1. (E) Correlations between timing variability and error variability (blue circles) and timing tolerance and error variability (red squares) on Day 11.
Figure 3.3. Experiment 2 Results. (A) Experimental task with changed target location. (B) Example arm trajectories in solution space from Day 1 (red) and Day 11 (blue). White circles represent the points of release. (C) Error variability and task success across practice. (D) Timing variability and timing tolerance across practice. Shaded regions represent ± 1 standard error of the mean. (E) Correlations between timing variability and error variability (blue circles) and timing tolerance and error variability (red squares) on Day 3 of practice.
Table 3.1A. Correlations of timing tolerance and error variability in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
<th>Day 6</th>
<th>Day 7</th>
<th>Day 8</th>
<th>Day 9</th>
<th>Day 10</th>
<th>Day 11</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>r</em></td>
<td>0.247</td>
<td>-0.580</td>
<td>-0.813</td>
<td>-0.797</td>
<td>-0.763</td>
<td>-0.962</td>
<td>-0.829</td>
<td>-0.847</td>
<td>-0.737</td>
<td>-0.799</td>
<td>-0.693</td>
</tr>
<tr>
<td><em>p</em></td>
<td>0.464</td>
<td>0.048</td>
<td>0.001</td>
<td>0.002</td>
<td>0.004</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.006</td>
<td>0.002</td>
<td>0.013</td>
</tr>
</tbody>
</table>

The shaded boxes indicate significant correlations (*p*<.05)

Table 3.1B. Correlations of timing variability and error variability in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
<th>Day 6</th>
<th>Day 7</th>
<th>Day 8</th>
<th>Day 9</th>
<th>Day 10</th>
<th>Day 11</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>r</em></td>
<td>0.247</td>
<td>-0.580</td>
<td>-0.813</td>
<td>-0.797</td>
<td>-0.763</td>
<td>-0.962</td>
<td>-0.829</td>
<td>-0.847</td>
<td>-0.737</td>
<td>-0.799</td>
<td>-0.693</td>
</tr>
<tr>
<td><em>p</em></td>
<td>0.464</td>
<td>0.048</td>
<td>0.001</td>
<td>0.002</td>
<td>0.004</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.006</td>
<td>0.002</td>
<td>0.013</td>
</tr>
</tbody>
</table>

The shaded boxes indicate significant correlations (*p*<.05)
4. Persistence of Reduced Neuromotor Noise in Long-Term Motor Skill Learning

Reward has been recognized as one of the most critical mediators of motor learning for over a century (Sternad & Körding, 2015; Thorndike, 1927). The objective of this study is to determine if reward can be used not only to decrease task errors, but also to attenuate neuromotor noise. This ever-present noise is the result of random fluctuations from all levels of the sensorimotor system and is a significant contributor to the variability observed in motor behavior. Typically, this intrinsic noise is assumed to be immune to practice and considered the limiting factor of performance. To allow for changes in this seemingly resistant feature, the present study examined extended practice. Further, retention was tested as only persistent changes have any value for rehabilitation. Demonstrating that neuromotor noise is reducible through practice would encourage attempts to diminish the exaggerated motor variability in some patient populations. This experiment used the throwing task as variability and reward could be easily manipulated.

I contributed to the experimental design, data collection, data analysis, statistical testing, modeling, and interpretation of results.

4.1 Abstract

It is well documented that variability in motor performance decreases with practice. Yet, the neural and computational mechanisms that underlie this decline, particularly during long-term practice, are only little understood. Decreasing variability is generally ascribed to error corrections from one trial to the next. However, the ubiquitous noise from all levels of the sensorimotor system, is also a significant contributor to overt
variability. While neuromotor noise is typically assumed and modeled as immune to practice, the current study challenged this notion. We investigated the long-term practice of a novel motor skill to test if neuromotor noise is attenuated, specifically when aided by reward. Results showed that both reward and self-guided practice over 11 days improved behavior by decreasing the noise processes, rather than effective error corrections. When increasing the challenge for reward, subjects reduced noise even further. Importantly, when task demands were relaxed again, this reduced level of noise persisted for five days. A stochastic learning model replicated both the attenuation and persistence of noise by scaling the noise amplitude as a function of reward. More insight into variability and intrinsic noise and its malleability has implications for training and rehabilitation interventions.

4.2 Introduction

Acquiring a new motor skill requires long hours of practice and patience, regardless of whether the goal is to play for the Boston Bruins or to merely make the high school hockey team. The same patience is required when recovering from brain injury such as stroke, where many months must be dedicated to relearn seemingly basic motor skills (Jørgensen et al., 1995a). And yet, when motor learning is studied in the laboratory, for practical reasons rarely do practice sessions exceed a single hour. Despite several experimental documentations of long-term practice and retention, little is known about what neural or computational mechanisms bring about the fine-tuning and persistence of the acquired skill (Nourrit-Lucas, Zelic, Deschamps, Hilpron, & Delignières, 2013; Park et al., 2013; Park & Sternad, 2015). The present study examined (1) what processes may
underlie the long-term improvements and retention, and (2) how can such honing of skill be enhanced.

Skill learning is marked by decrements in error and variability, starting with relatively rapid changes and followed by subtle tuning that continues over weeks, if not years of practice. In short-term motor adaptation studies, reduction in error and variability with practice has been attributed to the amount of error corrected from one trial to the next, i.e. the correction gain (Herzfeld, Vaswani, Marko, & Shadmehr, 2014; Smith, Ghazizadeh, & Shadmehr, 2006; Thoroughman & Shadmehr, 2000). However, it is an open question how variability continues to decrease once the error correction gain has been optimized and errors become increasingly smaller (van Beers, van der Meer, & Veerman, 2013).

In addition to explicit errors, random fluctuations, often ascribed to neuromotor noise, contribute to the overt variability in motor behavior (Izawa & Shadmehr, 2011; van Beers, 2009). This noise arises from the many interconnected layers of deterministic and stochastic processes throughout the neuromotor system, such as from variations in synaptic transmission and motor unit recruitment (Faisal et al., 2008). Even though prior studies have shown that the magnitude of variability and the implied noise can depend on movement speed (Fitts, 1954; Woodworth, 1899) and signal strength (Harris & Wolpert, 1998), this physiological noise is commonly viewed and modeled as immune to practice (van Beers et al., 2013; van Beers, 2009). The present study challenged this assumption. We hypothesized that in long-term skill learning, the amplitude of the noise decreases to attenuate overt variability.
If noise can decrease, then augmented feedback known to facilitate learning should be able to enhance this effect. Reward or “satisfying” consequences, recognized as a mediator for learning for over a century (Thorndike, 1927), has received renewed attention with supporting evidence in motor adaptation (M. Abe et al., 2011; Galea et al., 2015; Izawa & Shadmehr, 2011; Nikooyan & Ahmed, 2015; Shmuelof, Krakauer, & Mazzoni, 2012). Grounded in this rich history and recent evidence on reinforcement learning, we used reward to determine the extent to which neuromotor noise could be diminished.

We performed two experiments to test the overall hypothesis that novel skill learning proceeds by reducing neuromotor noise over long-term practice. In each experiment, we assessed learning of a virtual throwing task over 11 daily sessions. In Experiment 1, we hypothesized that reward leads to faster learning and better performance, both in error and variability, compared to self-guided learning (Hypothesis 1). Further, we expected that with extended practice, the fine-tuning of skill is achieved by decreasing the amplitude of noise (Hypothesis 2). Initial experimental results showed that while reward accelerated learning, both the reward and the self-guided control group reached the same level of performance and noise after 11 practice days. Hence, in Experiment 2, we hypothesized that increasing the reward incentive would lead subjects to decrease their noise amplitude further (Hypothesis 3). We additionally expected that the low level of noise persisted for five days after removing the increased incentive, indicating long-term retention of the skill (Hypothesis 4). A simple iterative model with a noise source that depended on reward was used to explain these experimental findings including retention.
4.3 Materials and Methods

4.3.1 Participants

18 healthy, right-handed students (9 females and 9 males, mean age 25.1 ± 2.3 years) from Northeastern University took part in the two experiments. None had any prior experience with the experimental task. Each participant performed 240 throws per day of a virtual throwing task for 11 days, resulting in a total of 2,640 throws. All participants gave informed written consent before the experiment and received monetary compensation upon completion of all 11 sessions. The experimental protocol was executed in compliance with the Institutional Review Board of Northeastern University.

4.3.2 Experimental Task and Apparatus

The experimental task was modeled after the British pub game skittles. In skittles, the player throws a ball tethered to a center post to hit a target skittle at the far side of the post (Figure 4.1A). In the experimental task, the participant manipulated a horizontal lever arm with a single-joint rotation about the elbow to throw a ball to hit a target in a 2D virtual environment (Figure 4.1B, see also Cohen and Sternad, 2009).

A top-down view of the skittles workspace was displayed on a rear projection screen approximately 60cm in front of the participant. Figure 4.1C shows the virtual scene, which consisted of the post (radius = .25m) centered at the origin, the target (radius = .05m); located 105.5cm above and 5cm to the right of the origin, the ball (.05m), and the lever arm (length = .4m) with the axle located 1.5m below the origin. The positions and size of each object was defined in a virtual coordinate system, and then scaled to project the graphics on the large screen. Participants moved a real lever arm to control the angle of the virtual arm. The participant’s forearm was secured to the
horizontal manipulandum that rotated around an axle centered at the elbow joint. A potentiometer (Vishay Spectrol, Ontario, Canada) at the axle continuously sampled the angular position of manipulandum at ~700Hz using a DT300 data acquisition board (Data Translation, Inc., Marlboro, MA) into a custom C++ program which ran the simulation.

Participants were instructed to release the ball such that it travelled through the center of the target without hitting the post. At the start of each throw, the participant grasped a wooden ball affixed to the distal end of manipulandum and closed a contact switch (Trossen Robotics, Westchester, IL) with his or her index finger. This attached the virtual ball to the end of the virtual arm. The participant then moved his or her arm and triggered the ball release by extending the index finger to open the switch. Upon release, all participants saw the ball traverse an elliptical path in the virtual scene for 1.5s as determined by the angle and velocity of the manipulandum at the moment of release. After the ball stopped, a small portion of the ball’s trajectory close to the target center was presented as additional visual feedback. Error for a given throw was defined as the minimum distance between the ball path and the target center (Figure 4.1D). In the experimental conditions where subjects were rewarded for successful throws, the target turned from yellow to green when a throw resulted in an error smaller than the defined reward threshold, indicating a successful hit to the subject.

4.3.3 Task Model

A two-dimensional model where the ball was attached to the origin by two orthogonal massless springs determined the path of the ball upon release. Due to restoring forces proportional to the distance of the ball from the origin and the velocity imparted at release, the ball was accelerated around the origin traversing an elliptical trajectory. The equations for ball position in the x- and y-directions at time $t$ were:
\[ x(t) = A_x \sin(\omega t + \varphi_x) e^{-\frac{t}{\tau}} \]

\[ y(t) = A_y \sin(\omega t + \varphi_y) e^{-\frac{t}{\tau}} \]

The frequency \( \omega \) denotes the natural frequency of the system, and relaxation time \( \tau \) was used to introduce damping to approximate realistic behavior. The amplitudes \( A_x \) and \( A_y \) and the phases \( \varphi_x \) and \( \varphi_y \) were calculated from the angular position and velocity of the ball at release based on recorded movement of the manipulandum. For the two experiments, \( \omega = 3.16 \text{rad/s} \) and \( \tau = 20 \text{s} \).

Since the two execution variables, angle and velocity at ball release, fully determined the ball trajectory, they also fully determined the result variable, error. The result space in Figure 4.1E illustrates this functional relationship: for each point in execution space, defined by release angle and velocity, the error for that throw is depicted by color: green defines successful (i.e. rewarded) solutions, yellow to black shades show increasing error. Perfect hits of the target (zero error) defined the solution manifold. The target location chosen for this study simplified the task such that release velocity did not affect the result; only the arm angle at ball release defined performance success (Figure 4.1E). This task configuration made the task faster to learn, allowing the focus of investigation on the refinement of motor skill, rather than on the initial exploratory stage. Note that different target locations lead to very different solutions manifolds (Sternad et al., 2014).

4.3.4 Experiment 1: Self-Guided Versus Reward Learning

Experiment 1 compared two groups that practiced with different conditions: the reward group (N=6) received reward in the form of a color change when the error was
below a threshold. The reward threshold was set to 1.1cm. The self-guided group (N=6) practiced without receiving any reward for successful trials.

4.3.5 Experiment 2: Changing Reward Incentive During Learning

Experiment 2 examined how the threshold for the reward signal influenced the performance. Hence, a changing-reward group of participants (N=6) practiced the task with different reward thresholds throughout practice. In the baseline block (Days 1-3), the reward threshold was 1.1cm as in Experiment 1; in the manipulation block (Days 4-6), the threshold was changed to 0.65cm, making it more difficult to receive a reward; in the persistence block (Days 7-11), the initial reward threshold of 1.1cm was reinstated, making reward again easier to obtain. To assess the influence of the reward threshold, performance of the changing-reward group was compared to that of the reward group from Experiment 1.

4.3.6 Dependent Measures

The dependent measures were analyzed for the 240 throws on each day of practice. The median and interquartile range (IQR) of the dependent measures were used because the Shapiro-Wilk tests revealed that the distributions of each measure were not normal on all days (Shapiro & Wilk, 1965). For all experimental groups, overall task performance was measured by the median error on each practice day. Because the error was unsigned, the median error on each practice day reflected the combined effects of bias and variability in performance. For the two reward groups, task performance was also evaluated by the percentage of rewarded throws (errors smaller than the reward threshold) of each practice day.

To characterize variability, error was not a suitable measure because the errors in this task were non-negative and therefore lower-bounded. Thus, the release angles were
used instead as they fully determined the error due to the specific solution manifold. As Figure 4.1E shows, with this target location, the solution manifold was oriented vertically, meaning that the success of the throw was independent of velocity. Therefore, the variability was quantified by the interquartile range, IQR, of release angles. To determine if reductions in release angle variability were possibly the result of slowed movements, the median release velocity on each day was analyzed.

In addition to this distributional analysis for each day, the temporal structure of the trial-by-trial sequence was also examined by calculating the autocorrelation coefficient of successive release angles. Specifically, we obtained the lag-1 coefficient (ACF-1) to assess the presence of noise in the observed variability (van Beers et al., 2013). ACF-1 values between -1 and 0 indicate that successive observations were negatively correlated, also known as anti-persistence; ACF-1 values between 0 and 1 indicate persistence, or positive correlation between successive observations. An ACF-1 value of zero indicated uncorrelated, white noise. In addition, detrended fluctuation analysis (DFA) was applied to the time series of release angles to identify temporal dependencies beyond lag-1 (M. O. Abe & Sternad, 2013; Dingwell & Cusumano, 2010; Peng et al., 1994). Importantly, DFA can identify persistence and anti-persistence without exact sequential dependencies. The scaling index quantifies the long-range correlations of the time series. Here it was applied to complement the results of ACF-1. Scaling indices above and below .5 indicate correlation and anti-correlation, respectively; a scaling index of 0.5 indicates unstructured noise.

4.3.7 Statistical Analyses

In each experiment, the dependent measures were analyzed using two-way repeated measures ANOVAs (group × day). Group was a between-participants factor and day was
within-participants factor. Relevant planned comparisons using paired \( t \)-tests were made to further probe the within-participant effects; independent sample \( t \)-tests investigated group effects. The significance level was set at \( \alpha = 0.05 \) for all statistical tests.

4.4 Results

4.4.1 Reward Accelerates Error Reduction, but After Extended Practice, Self-Guided Practice Reaches the Same Level of Performance

Experiment 1 tested \emph{Hypothesis 1} stating that subjects who were incentivized by reward would reduce both their error and variability faster and to a lower level (\emph{Hypothesis 1}). Subjects in the reward group (\( N=6 \)) received a visual signal of success after each throw when the ball trajectory passed through the target with an error was less than 1.1 cm (Figure 4.1D). The self-guided group (\( N=6 \)) performed the task without any reward signal and the target remained yellow in all cases.

As expected, subjects in both groups reduced their median error over 11 days of practice, as seen in a significant effect of day (\( F(10,100) = 9.57, p = .002 \)) (Figure 4.2A). The reward group also increased their percentage of rewarded throws from an initial average of 69% on Day 1 to 84% by Day 11. Subjects in the reward group performed with lower median error than the self-guided group, indicated by a significant group effect (\( F(1,10) = 15.62, p = .003 \)). The interaction was also significant (\( F(10,100) = 4.34, p = .034 \)) because the self-guided group began with greater median error on the first two days (\( p = .001 \) and \( p = .0046 \)), but approached the reward group on most of the following days (see Figure 4.2A). This interaction suggests that reward accelerates error reduction as
hypothesized. However, with sufficient practice, the self-guided group reached the same level of error, which ran counter the *Hypothesis 1*.

### 4.4.2 Self-Guided and Reward Learning Have Similar Variability

To quantify the precision of performance, the next focus of analysis was on variability of throws within each practice day. As expected, all subjects significantly reduced their variability, as measured by the IQR of release angles, over days ($F_{(10,100)}=11.86, \ p<.001$, Figure 4.2B). However, the two groups did not differ significantly from each other ($F_{(1,10)}=3.89, \ p=.077$): while subjects in the self-guided group started with significantly greater variability on Day 1 ($p=.014$), this group difference disappeared over the remaining 10 days ($ps>.075$, Figure 4.2C). The Group x Day interaction was non-significant ($F_{(10,100)}=2.73, \ p=.070$). This result also ran counter to *Hypothesis 1*.

### 4.4.3 Variability in Both Self-Guided and Reward Learning is Random

Decreases in error and variability over practice are commonly ascribed to error corrections, particularly in motor adaptation paradigms where error-based learning predominates. However, random fluctuations resulting from intrinsic neuromotor noise also affect these performance measures (M. O. Abe & Sternad, 2013; van Beers et al., 2013). *Hypothesis 2* stated that to improve in task performance over long-term practice, noise amplitude has to decrease.

Initially, the average ACF-1 calculated for each daily session was positive in all subjects of both groups, indicating weak persistence across successive throws (Day 1: $M=.26, SD=.16$). Over sessions, however, ACF-1 significantly decreased and approached zero ($F_{(10,100)}=5.30, \ p=.001$, Figure 4.2D). From Day 5 onwards, the average ACF-1 of all subjects and both groups was within the 95% confidence interval of uncorrelated white
noise. Neither the group main effect ($F_{(1,10)}=0.69, p=0.43$), nor the interaction were significant ($F_{(10,100)}=1.82, p=0.14$). To further test Hypothesis 2, detrended fluctuation analysis was applied over the same data. The scaling index similarly approached 0.5, which is the value for unstructured noise, corroborating the findings of the autocorrelation results (see Figure 4.3A-B). These results indicate that by the end of practice, neuromotor noise predominated, consistent with Hypothesis 2.

4.4.4 Reduced Variability and Noise Was Not the Result of Reduced Movement Speed

While the lack of temporal structure supported the expectation that neuromotor noise attenuated with practice, alterative explanations for reducing overt variability have been proposed. For instance, a common strategy in motor performance is to lower movement speed in return for accuracy (Fitts, 1954). Examining the change in velocity at which participants released the ball did not reveal significant decrements in median release velocity across practice days ($F_{(1,100)}=1.73, p=0.19$), nor group differences ($F_{(1,10)}=2.06, p=0.18$), or an interaction ($F_{(1,100)}=0.73, p=0.52$) (Figure 4.2E). Hence, there is no sign that a speed-accuracy trade-off could be responsible for the observed improvements in accuracy.

A second widely accepted mechanism underlying overt variability is that noise is dependent on signal strength (Harris & Wolpert, 1998). This has been illustrated in an optimal control model and experimentally shown in isometric force production, where the increased variability at higher forces has been ascribed to signal-dependent noise (Jones, Hamilton, & Wolpert, 2002). In dynamic tasks, high velocity requires fast muscle contractions; hence, velocity is the dynamic correlate of high force levels and faster movements are expected to be associated with higher noise. The fact that release velocity
did not change with practice therefore also ruled out that subjects exploited this possibility for noise reduction (Sternad et al., 2011). This lack of support for alternative explanations indirectly corroborates that the reduced variability was the result of reduced noise.

4.4.5 Neuromotor Noise is Reduced Further by Changing Reward

In Experiment 1 reward led to faster reduction of error, but not to lower error or variability by the end of practice, counter to Hypothesis 1. However, in support of Hypothesis 2, noise processes dominated the observed variability in the sequence of throws by the end of practice. This second finding may imply that subjects have reached their physiological limit. Alternatively, subjects may only have reached a point of diminishing returns, where any further reduction of noise was not worth the effort, especially given their high success rate (84% of trials were successful by the end of practice). To discern between these two possible explanations, Experiment 2 manipulated the reward threshold to make it harder to achieve reward. Hypothesis 3 posited that the higher demands would elicit further decreases in the amplitude of noise. As a direct test, Experiment 2 first presented the same threshold for reward as in Experiment 1 over Days 1 to 3, but changed the threshold over Days 4 to 6 to require more accurate throws to receive reward (Figure 4.4A). We expected that the noise amplitude would be lowered.

To further test that practice induced persistent effects, the tighter task demands were relaxed again over Days 7 to 11. While in motor adaptation, the acquired behavior invariably returns to the original behavior after removing the external perturbation, the hallmark of skill learning is that the learned behavior persists for a long time, indicative of neuroplasticity. Hence, to test that practice induced persistent effects, the initial
threshold was reinstated on Day 7 and practiced for 5 days until Day 11 (Figure 4.4A).

Hypothesis 4 stated that subjects would retain their acquired low level of noise.

On Days 1-3, subjects in both groups significantly improved task performance, marked by an increasing percentage of successful trials ($F_{(2,20)}=17.70$, $p<.001$, Figure 4.4B). Corroborating results of Experiment 1, they both decreased median error ($F_{(2,20)}=10.035$, $p=.001$, 5.4C), and release angle variability ($F_{(2,20)}=4.85$, $p=.039$, Figure 4.4D). The temporal structure in release angle variability as measured by ACF-1 did not show any significant change across days ($F_{(2,20)}=2.76$, $p=.088$, Figure 4.4E). The average ACF-1 of all subjects was within the 95% confidence interval of uncorrelated white noise on Days 2 and 3, suggesting that variability was due to noise on these days. The findings of autocorrelation analysis were also corroborated by the results of the DFA (see Figure 4.3B-C). As expected, there was no significant effect of group or Group x Day interaction on any of the dependent measures ($F$s<1.8), corroborating the results of Experiment 1.

Although variability in error and release angle decreased, release velocity did not change as a function of practice. In fact, release velocity increased over the three days ($F_{(2,20)}=9.66$, $p=.003$, Figure 4.4F). As in Experiment 1, this suggested that variability did not decline as a result of slower movement speeds as expected from Fitt’s Law.

During Days 4-6, subjects in the changing-reward group were exposed to an altered threshold of 0.65cm to enhance the task challenge; the threshold remained at 1.1cm for the reward group. Note that the visual size of the target was unchanged (Figure 4.4A). As expected, the success rate of the changing-reward group dropped significantly compared to the reward group ($F_{(1,10)}=31.92$, $p<.001$, Figure 4.4B), confirming that task success was indeed more difficult for these subjects. Median error and release angle variability
showed trends suggesting that the changing-reward group improved faster than the reward group as hypothesized, although these interactions did not reach statistical significance ($p$s<.17). There were no significant effects or interactions in any other dependent measures ($p$s>.05).

4.4.6 Persistence of Reduced Noise

While subjects only gradually responded to the change in the reward threshold, a significant group difference did emerge on Day 7, when the threshold was restored to the original value. Even though both groups practiced with the larger reward threshold for the remaining 5 days, the changing-reward group continued with a significantly higher success rate ($F_{(1,10)}=19.055$, $p=.001$, Figure 4.4B). This improved performance resulted from significantly lower error ($F_{(1,10)}=19.97$, $p=.001$, Figure 4.4C) and variability ($F_{(1,10)}=10.019$, $p=.01$, Figure 4.4D). These results were consistent with Hypothesis 4.

The temporal structure of variability did not differ between the two groups ($p$s>.05) and the average ACF-1 for all subjects in both groups was within the 95% confidence interval of uncorrelated white noise for all 5 days (Figure 4.4E). No significant effect of day or interaction was observed in any of the measures ($p$s>.05). There was no group difference in median release velocity ($F_{(1,10)}=.66$, $p=.44$), which again ruled out that speed-accuracy trade-off or signal strength may have lowered the overt variability or noise.

These behavioral results suggest that it was physiologically possible for subjects in the changing-reward group to decrease the noise in their motor behavior and achieve better task performance.
4.4.7 A Simple Model with Time-Varying Gain on Noise Reproduces Empirical Results

To illustrate how the change in success requirements could influence motor variability via its two essential components error correction and noise, we used an extremely simple, yet prevalent iterative learning model. The model describes how the central nervous system might correct movement errors on a trial-by-trial basis:

\[ x(n+1) = x(n) - Be(n) + \xi(n) \]  
\[ e(n) = x(n) - x^* \]

(Eq. 1)  
(Eq. 2)

where \( x(n) \) is the motor output at trial \( n \), and \( e(n) \) is the error at trial \( n \), defined as the difference between the actual and desired motor output \( x^* \). The parameter \( B \) defines the fraction of error corrected, or error correction gain, \( \xi \) is an additive white Gaussian noise, \( \xi \sim N(0, \sigma^2) \). This model in its current form and with slight variations has been successful in simulating a range of phenomena associated with motor adaptation (Herzfeld et al., 2014; Smith et al., 2006).

On the basis of this model, how can variability be decreased? In principle, there are two options: either increase the error correction gain or decrease the amplitude of the noise. To determine if an increase in error correction gain could explain the decrease in variability, we simulated data using Eqs.1 and 2 with five different \( B \) values, while keeping the noise amplitude \( \xi \) constant. The motor output \( x(n) \) in the model represented the release angle for a given trial \( n \); \( x^* \) was the optimal release angle that resulted in zero error; \( e(n) \) represented the error in release angle. Although error in release angle is different from the error measure used in the experiments, these two error measures are tightly related. As shown in Figure 4.1E, the error increases with the difference between the release and the optimal angle, though not exactly linearly.
For each value of $B$ (.1, .3, .5, .7, .9), the model was simulated for 2,640 trials to match the total number of trials performed by the human subjects. The results were analyzed in bins of 240 trials representing “days” to match the analysis of the experimental data. The optimal angle, $x^*$, for this target location was 82.44° (Figure 4.1E). The average median release angle on Day 1 from subjects in the reward and changing-reward groups was used as initial values of $x$, such that $x(1) = 83°$, and the average IQR on Day 1 from all subjects in the reward and changing-reward groups was used to set the variance of the additive white Gaussian noise, such that $\sigma^2 = 8.7°$.

Figure 4.5A shows that the amplitude of variability in release angle, is lowest for higher error correction gains ($B \sim 1$). Importantly, the amplitude of variability remained relatively constant over simulated days of practice regardless of the value of $B$. Hence, to decrease variability through the error correction gain alone, $B$ must increase over trials, i.e. it must become a function of trials, $B(n)$.

However, Figure 4.5B also demonstrates how different $B$ values produced different temporal structure in the variability: the higher the error correction gain, the closer to zero the ACF-1 value becomes (M. O. Abe & Sternad, 2013; van Beers et al., 2013; van Beers, 2009). In Experiment 2, the ACF-1 value was within the range of white noise from day 2 onwards, suggesting that the error correction gain $B$ was already high (~.90) at the beginning of practice. Furthermore, the ACF-1 did not significantly change over days. Thus, it is unlikely that the reduced variability observed in the experimental data was achieved by an increasing error correction gain.

This suggests that the noise term had to be scaled. We introduced a gain factor $a(n)$ to modulate the amplitude of noise in the model as follows:
\[ x(n+1) = x(n) - Be(n) + a(n)\xi(n). \]  

(Eq. 3)

Next, we defined how this noise gain changed over practice. For both reward groups, the results of Experiment 2 showed a clear relationship between the reward threshold and the magnitude of random variability: subjects modulated noise to increase reward. This finding concurs with previous results in a motor adaptation task by Izawa and Shadmehr who demonstrated that reward could be used to modulate noise (Izawa & Shadmehr, 2011). However, in their study subjects amplified noise to explore solutions that yielded a higher reward. In the present study, subjects were already centered on the optimal solution (Figure 4.1E). To exploit this correct solution and maximize reward, it was advantageous to decrease noise.

Therefore, the following rule for updating the noise scaling factor \( a(n) \) was added

\[ a(n+1) = (1 - 1_{e(n)>T}A)a(n) \]  

(Eq. 4)

where \( e(n)>T \) denotes the condition that error is greater than the reward threshold, \( T \). The term \( 1_{e(n)>T} \) denotes an indicator function that equals 1 if absolute value of error \( e(n) \) is greater than the threshold \( T \) and no reward is given. The function is set to 0 if reward is received. \( A \) denotes the rate at which the noise scaling factor decreases. According to this update rule, \( a(n) \) is decreased whenever no reward is obtained. Figure 4.5C shows results from the iterative learning model together with the results of Experiment 2. Note that in the experiment, the reward threshold for success was based on error. For the simulation, the error threshold was transformed into a corresponding angle threshold. Assuming a release velocity of 500m/s, a 1.1cm error threshold corresponded to an 8.9° angle threshold, and a .65cm error threshold corresponded to a 4.4° angle threshold. The value of \( B \) was set to 0.9 and \( A \) was set to 0.0015 to match the temporal and spatial
characteristics of the variability observed in the experimental data, and \( a(1) \) was set to 1. The model replicated the slow decline in release angle variability after the change in threshold on Days 4-6 in the changing-reward group. It also shows that the release angle variability is maintained after reinstating the initial reward threshold.

Note that the self-guided group also scaled down the noise over practice, eventually reaching the level of the constant reward group. However, it is unclear from the experimental protocol what conditions were responsible for the decrease \( a(n) \) from one trial to the next. While self-guided learning is evidently ubiquitous in the real world, studying this “uncontrolled” form of learning with a model-based approach remains a challenge.

### 4.5 Discussion

Mastering a motor skill can take a lifetime, however most motor learning and adaptation studies have examined only on the initial phases of learning. Contrastingly, this study focused on the later fine-tuning process of skill learning, where only subtle changes occur during many hours of practice. The overall hypothesis was that the characteristic decline in variability with practice is due to an attenuation of the magnitude of neuromotor noise, rather than increased error corrections. To enhance this slow tuning process, subjects were given reward for additional feedback and motivation. We further hypothesized that with reward, this neuromotor noise would not only decrease, but also persist for an extended time, indicative of lasting plastic changes in the nervous system.
4.5.1 Frequent Reward Accelerates Learning but Does Not Lead to Better Performance

According to many motor learning textbooks, augmented feedback is arguably the single most important variable for motor learning (Salmoni et al., 1984; Schmidt & Lee, 2011). However, the experimental results qualified this straightforward expectation and highlighted that while augmented feedback did promote learning, in accord with Hypothesis 1, it failed to provide an advantage over self-guided practice in the longer run. Hence, internal feedback mechanisms via visual and proprioceptive information were equally effective. This finding resonates with numerous examples in real-world scenarios and experimental settings, where learning proceeds without controlled external feedback and reward (Park et al., 2013; Park & Sternad, 2015).

However, results also showed that when the reward became sufficiently challenging and less frequent, it could provide an advantage over self-guided practice. Apparently, subjects had not reached their limit as increased accuracy requirements for the success signal could elicit further improvements (Hypothesis 3). One reason may be that the cost to lower noise may have been too high, consistent with the assertion of Manohar et al. (2015). Alternatively, there may also be advantages to variability in motor behavior, such as supporting exploration. Recently, Wu et al. (2014) provided renewed evidence that initial variability can be beneficial for exploring the solution space to find the behavior that receives reward. Yet, even after the initial exploration, it is possible that subjects maintain a certain level of noise in order to gain information about the task (Kaelbling, Littman, & Moore, 1996). Frequent reward may have discouraged exploration and instead prioritized exploitation. It has been also hypothesized that it is not the reward per se that drives learning, but rather the mismatch between the expectation
and receipt of a reward (Tobler, O’Doherty, Dolan, & Schultz, 2006). The fact that the changing-reward group did not receive reward for trials that were previously rewarded may explain why this more difficult reward led subjects to further reduce their neuromotor noise.

4.5.2 Neural Mechanisms to Reduce Neuromotor Noise

While it is not surprising that variability decreased under tighter task demands, it is non-trivial that it was the decrease in the amplitude of neuromotor noise, not an increase in error correction that reduced the overt variability (Hypothesis 2 and 3). How might the central nervous system reduce not only overt, but possibly also physiological noise? One potential mechanism is to increase antagonistic co-activation (Gribble, Mullin, Cothros, & Mattar, 2003; Selen, Beek, & van Dieën, 2005), although antagonistic co-activation tends to decrease with practice (Thoroughman & Shadmehr, 1999). Lower signal-dependent noise has been ruled out as a candidate by the experimental results that showed no decrease in velocity (Harris & Wolpert, 1998). Another conjecture is derived from studies on the effect of neuromodulators on motor neuron excitability, such as serotonin (Theiss & Heckman, 2005; K. Wei et al., 2014), norepinephrine (Theiss & Heckman, 2005), and dopamine (Kroener, Chandler, Phillips, & Seamans, 2009). Animal studies provided intriguing evidence that the descending drive to muscle contractions is gain-controlled to modulate the required output force. One study on humans specifically showed that force variability increased after the brainstem–spinal cord neuromodulatory system was up-regulated (K. Wei et al., 2014). The complex interplay of neuromodulators can excite or inhibit spinal cord excitability and thereby may match precision demands in motor tasks. Finally, Picard and colleagues demonstrated in skilled performance of monkeys that years of practice resulted in more efficient generation of
neuronal activity in M1 (Picard, Matsuzaka, & Strick, 2013). It is feasible that such a decrease in metabolic activity could also result in lowered neuromotor noise. Alternatively, the variability in neuronal activity may be shaped and reduced in neural space and in a way that leads to more consistent performance (Sadtler et al., 2014). Evidently, more research is needed to solidify these conjectures.

4.5.3 Long-Term Persistence Indicative of Long-Term Neuroplasticity

Experiment 2 not only tested whether modulation of reward threshold influenced learning, but also whether this effect showed persistence over the five days after the reward criterion was relaxed. While Hypothesis 4 stated that noise would remain suppressed, the experimental result was nevertheless stronger than expected. At first blush, it appeared similarly reasonable that random fluctuations might have increased after the task constraints were alleviated, as previous studies provided ample documentation that humans shape their variability rather than only reduce it (Chu et al., 2013; Cohen & Sternad, 2009; Müller & Sternad, 2004b; Sternad et al., 2014). However, having suppressed the magnitude of random fluctuations, subjects may not have sensed that relaxing their performance variability was possible as there was no verbal or visual indication. Alternatively, and following Hypothesis 4, subjects may indeed have permanently altered their performance, due to lasting changes in the nervous system.

Note that this persistence differs from savings, which is one sign of neuroplastic effects in motor adaptation; savings refers to the reduced time when exposed to the same perturbation more than once. In adaptation, subjects experience large errors when the perturbation is removed and, consequently, the original behavior is reinstated to perform the task successfully. In the present study, the acquired behavior was the correct behavior in all conditions. Thus, the learned behavior with reduced noise could persist without
penalty after the manipulation was removed. This result underscores that interventions to enhance long-term behavior should use conditions that enhance a behavior that remains essentially identical, even when the intervention is removed (Huber & Sternad, 2015; Reisman, Wityk, Silver, & Bastian, 2009).

4.5.4 Considerations and Limitations of Modeling Skill Learning

To illustrate the interplay of error, correction, and noise in trial-to-trial learning, an extremely simple recursive model generated results that matched the behavioral data observed in Experiment 2. The model combined the usual error correction gain with an additive noise term and a time-varying noise gain that was altered by a simple indicator function representing the threshold for reward. The model successfully simulated the decrease in variability and the persistence of the low level of noise. Interestingly, it also captured the delayed response to the changed reward that only became significant after three days of practice. Note that the present model included only a single additive noise source and no multiplicative noise source. Other research has shown that two or more noise sources are required to account for observed temporal structure in the data (M. O. Abe & Sternad, 2013; van Beers, 2009).

Despite its simplicity, the simulation demonstrated how noise might decrease in response to the change of reward feedback. The behavioral results of the self-guided group also showed a decrease in noise over practice. However, it remains unknown what drove these subjects to lower their noise, as there was no comparison group to test a hypothesized mechanism. One challenge with modeling self-guided, or unsupervised, skill learning is that by definition, there is no explicit or added performance feedback given during learning (Wolpert, Ghahramani, & Flanagan, 2001). Thus it is difficult to discern, and hence simulate, the exact information subjects are using to update their
performance from one trial to the next trial. Nevertheless, efforts to better understand this type of skill learning are needed because this is probably the most prevalent form of practice in the real world. As with any modeling attempt, behavioral and physiological data are needed to reveal the patterns of behavior when we acquire and retain complex motor skills.

Modeling long-term skill learning that addresses the ever-present stochastic components still remains an open challenge in the field of motor neuroscience. Neural networks present an alternative to explore the role of noise in skill acquisition, long-term retention, and generalization (Ajemian et al., 2013). However, stochastic models and system identification are still in their infant stages in motor neuroscience, despite the recognized critical role of noise in the nervous system.

4.6 Conclusion

Long-term skill learning and retention requires attention as it opens an interesting view on a ubiquitous but ill-understood element of the neuromotor system - noise. Our experimental results suggest that skill learning proceeds by reducing the amplitude of neuromotor noise. Reward enhanced this change, but self-guided practice was also effective. The intriguing finding was that this enhanced performance persisted for days, even after incentive was decreased. This study highlights the need to further understand neural and computational processes underlying long-term learning, especially as we continue to look for avenues to enhance skill learning and expedite rehabilitation.
4.7 Figures

Figure 4.1. Task description for the skill learning experiments. (A) Real skittles throwing task. The experimental task was modeled after the British pub game skittles. In skittles, the player throws a ball, tethered to the top of a post, around that post to hit a target skittle on the other side. (B-C) Experimental setup of virtual throwing task. Subjects manipulated a horizontal lever arm with a single-joint rotation about the elbow to throw a virtual ball in a 2D virtual environment as seen in c. As in the real game, they were instructed to throw the ball such that it travelled through the center of the target without hitting the post. (D) Error is defined as the minimum distance between the ball path and the target center. If the error is below the reward threshold, the subject receives a reward indicating a successful target hit; otherwise there was no added feedback. The reward is given in the form of a color change of the target from yellow to green. Subjects in the reward and changing-reward groups were instructed to achieve as many target hits as possible. Note that for subjects in the self-guided group the target always remained yellow, regardless of error. (E) Result space of modeled skittles task. The result space illustrates the functional relation between the two variables that determine the ball path, release angle and velocity, and the result variable, error: for each combination of release angle and velocity, error is depicted by color (darker shades indicate increasing error). The dots depict 60 throws (represented by their release angle-velocity pairs) of an example subject in the reward group during the first (magenta) and last (blue) days of practice.
Figure 4.2. Reward accelerates error reduction, but with sufficient practice, self-guided practice reaches a similar level of performance and noise. (A-E) Performance from subjects in the self-guided group is shown in blue, and performance from those in the reward group is shown in red. (A) Group comparison of error over 11 days of practice. The self-guided practice group had higher error for most of the practice days, however they converged to a similar error level as the reward group by the end of practice. (B) Comparison of variability between self-guided and reward groups. While the self-guided group had higher variability on the first day of practice, this difference disappeared for the remaining 10 days of practice. (C) Time series of release angles during the first four days and the last two days of practice of an example subject in the self-guided group (blue) and reward group (red). (D) After 4 days of practice, the mean lag-1 autocorrelation values (ACF-1) for both groups reveal no temporal structure in release angle. This indicates that the variability in release angle after extended practice was dominated by random noise. (E) Individual subjects did not decrease their release velocity as a strategy to decrease release angle variability. High inter-subject variability results from the fact that release velocity played very little role in determining task performance. All error bars represent the ±2 s.e.m. Asterisks indicate significance (*p < 0.05).
Autocorrelation and detrended fluctuation analyses consistently show that trial-to-trial fluctuations converged to a noise process. (A-C) Autocorrelation and detrended fluctuation analyses were conducted for each experimental group. ACF-1 values close to 0 and scaling exponent values of 0.5 indicate unstructured, random noise. Both analyses produced similar results. All error bars represent the ±2 s.e.m.

**Figure 4.3.** Autocorrelation and detrended fluctuation analyses consistently show that trial-to-trial fluctuations converged to a noise process. (A-C) Autocorrelation and detrended fluctuation analyses were conducted for each experimental group. ACF-1 values close to 0 and scaling exponent values of 0.5 indicate unstructured, random noise. Both analyses produced similar results. All error bars represent the ±2 s.e.m.
Figure 4.4. The reward condition that requires more accurate performance leads to lower variability that persists over 5 additional practice days. (A) The reward group, which is the same as in Experiment 1, practiced with a reward threshold of 1.1cm for all 11 days of practice. The changing-reward group practiced with a reward threshold of 1.1cm for the first 3 days of practice, followed by 3 days with a more challenging reward threshold of 0.65 cm and 5 days with the original threshold of 1.1cm. (B-E) Group comparisons of dependent measures. Performance from subjects in the reward group is shown in red, and performance from those in the changing-reward group is shown in green. The shaded region indicates the practice days where the reward threshold between the two groups differed. (B) The percentage of rewarded throws in each practice day was reduced with the more challenging threshold as expected for the changing-reward group. When the original reward threshold was reinstated, subjects in the changing-reward group were more successful than the reward group. (C-D) While group differences in error and variability did not arise during the reduced reward threshold as expected, they did emerge during the persistence period, as the changing-reward group performed the task with lower error and lower variability. (E) The average ACF-1 of all subjects was within the 95% confidence interval of uncorrelated white noise from the second day of practice onwards. (F) As in Experiment 1, individual subjects did not decrease their release velocity as a strategy to decrease release angle variability. All error bars represent the ±2 s.e.m. Asterisks indicate significance (*p < 0.05).
Figure 4.5. Simulation of decreased neuromotor noise as a function of binary reward reproduces empirical results. (A) Simulated data using the iterative model (Eqs. 1 and 2) with different $B$ values while keeping the noise amplitude constant. For each simulated value of $B$, variability measured by the interquartile range remained relatively constant over simulated days of practice. (B) Different $B$ values result in different temporal structure in the variability. For higher error correction gains, ACF-1 approaches zero. (C) Incorporating the update rule that decreases the noise amplitude as a function of reward (Eqs. 3 and 4) reproduced the empirical results of Experiment 2. Six simulations were conducted for each group. All error bars represent the ±2 s.e.m. The changed threshold gradually leads to a difference between the two groups. The rate at which the group difference arises is dependent on the rate at which the noise scaling factor $A$ decreases in (Eq. 3). For these simulations, the value of $A$ was chosen to best reproduce the empirical results of Experimental 2. The group difference is maintained after reinstating the initial threshold.
5. Implicit Guidance to Stable Performance in a Rhythmic Perceptual-Motor Skill

Current research on virtual and robotic rehabilitation use predominantly error- and reward-based feedback to expedite learning. These approaches have been effective in guiding learning of many simple tasks as demonstrated in the previous chapter. However, activities of daily living are generally redundant tasks that require additional and subtler guidance to help the learner identify the mapping between execution and task outcome. Thus, there is a need to further develop principles of implicit guidance that go beyond the explicit error- and reward- based approaches.

The goal of this study is to identify elements of performance in a redundant skill to guide learners towards desired dynamically stable and noise-tolerant solutions. Moreover, the guidance method should lead to learned skill that persists also when the assistance is removed. Patients should not remain dependent on the assistance used in therapy. Specifically, this research aims to develop interventions that accelerate learning of a rhythmic task in a virtual environment. While rhythmic movements are ubiquitous, ranging from locomotion to many forms of tool use, much of the research on guiding learning and adaptation focuses solely on discrete movements. As detailed below, control of rhythmic movements involves different principles, most notably dynamic stability, that affords an entry point into guidance. Therefore, this study employs the rhythmic ball bouncing task, where theoretical foundations for dynamic stability have been previously developed (Dijkstra et al., 2004; Schaal et al., 1996; Sternad, Duarte, et al., 2000; K. Wei et al., 2007; K. Wei, Dijkstra, & Sternad, 2008).
I contributed to the experimental design, theoretical derivation of the task manipulation, data collection, data analysis, statistical testing, and interpretation of results.

5.1 Abstract

Feedback information about error or reward is regarded essential to aid learners to acquire a perceptual-motor skill. Yet, simple error feedback does not suffice in guiding the learner towards the optimal solutions, when tasks have redundancy where the mapping between execution and performance outcome is unknown. The present study developed and tested a new means of implicitly guiding learners to acquire a perceptual-motor skill, rhythmically bouncing a ball on a racket. Due to its rhythmic nature, this task affords dynamically stable solutions that are resistant to small errors and noise, a strategy that is independent from simply reducing error. Based on the task model implemented in a virtual environment, a state-dependent manipulation was designed that shifted the range of ball-racket contacts that achieved to dynamically stable solutions. In two experiments, subjects practiced with this manipulation that guided them to impact the ball with more negative racket accelerations, the indicator for the strategy with dynamic stability. Subjects who practiced under normal conditions took longer time to acquire this skill, although error measures were identical between the control and experimental groups. Unlike in many other haptic guidance or adaptation studies, the experimental groups not only learned but also maintained the stable solution after the manipulation was removed. These results are a first demonstration that more subtle ways to guide the learner to better performance are needed to assist performance improvements, especially in tasks with redundancy, where error feedback may not be sufficient.
5.2 Introduction

From learning to walk or playing a drum set, to relearning to write after a stroke, humans acquire a vast collection of motor skills over their lifetime. Many strive to improve advanced motor skills, while others have to relearn basic skills, such as coordinating a knife and fork to eat. Consequently, much research has been dedicated to understand and facilitate motor learning and recovery. Following decades of research on operant conditioning or reinforcement learning in the 1940-60s, a host of subsequent studies adopted a more cognitive approach to enhance skill learning, many under the umbrella of schema theory or generalized motor programs (Adams, 1987; I. M. Bilodeau, 1966, 1969; Magill & Anderson, 2010; Newell, 1976; Schmidt & Lee, 2011; Schmidt, 1975). Despite many unresolved issues, one consensus from these studies was that presenting augmented (as opposed to only intrinsic) information about the outcome following the action, or knowledge of results, enhances performance and learning (Salmoni et al., 1984). Numerous experiments aimed to identify its optimal frequency, temporal delay, and precision of delivery that may lead to optimal performance, although few generalizable results arose. This quantitative terminal feedback was sometimes separated from knowledge of performance, defined as augmented kinematic and kinesthetic information about the movement itself (Newell & Walter, 1981). However, only relatively few studies were conducted to further detail its potential benefit, probably due to the lack of virtual technology.

More recently, error-based learning and feedback information have been studied in a discrete reaching paradigm requiring adaptation to external force fields and visuomotor rotations. Using virtual environments, several studies validated that presenting
continuous and terminal error information aids in learning the new visual-motor remapping or the compensation of external forces (e.g., Hinder et al. 2008; Shabbott and Sainburg 2010). Besides visual information about the endpoint effector, additional proprioceptive information was examined to support and accelerate adaptation (Scheidt, Conditt, Secco, & Mussa-Ivaldi, 2005; Scheidt, Lillis, & Emerson, 2010). More recently, the role of terminal reinforcement or reward scores and credit assignment has been highlighted (M. Abe et al., 2011; Galea et al., 2015; Wolpert et al., 2011). Characteristic to these studies is that adaptations to new conditions faded relatively quickly, when the perturbation was removed. Further, regardless of the error information, subjects adapted to new rotations or force fields in as few as 10 trials, and return to initial performance (after-effects) has been even shorter (Kitago, Ryan, Mazzoni, Krakauer, & Haith, 2013). These fast performance changes highlight that reaching and its transformations are easy to learn and present relatively little challenge. The correct solution is uniquely defined by the target location and the desired straight-line trajectory, which makes error information very effective (Scheidt et al., 2005, 2010).

In contrast to this adaptation of reaching movements, acquiring a novel skill takes considerably longer, from weeks to years, and presenting knowledge of results is not as straightforward and effective (Cesqui et al., 2012; Crossman, 1959; Park et al., 2013; Park & Sernad, 2015). Any athlete or coach knows that a single-numbered score, as for example given by judges, is not sufficient information to improve performance in a complex motor skill. After all, finding ways that shape behavior towards a desired goal is what makes or breaks a good coach. By definition, in a novel skill the relationship between motor execution and task outcome is unknown and the mapping is typically very
complex. Not only does the redundancy in the neuro-mechanical system afford infinitely many ways to produce the desired movement, the task itself can also be achieved in infinitely many ways and performance improvement may proceed in a non-monotonic fashion. The manifold of solutions in the result space is typically nonlinear, with some solutions more favorable than others (Berret, Chiovetto, Nori, & Pozzo, 2011; Campolo, Widjaja, Xu, Ang, & Burdet, 2013; Ganesh & Burdet, 2013; Sternad et al., 2011, 2014).

Specifically, some solutions tend to be more tolerant to perturbations and noise. Previous research by Sternad and colleagues highlighted this redundancy in a discrete throwing task, where the set of zero-error solutions described a nonlinear manifold (Müller & Sternad, 2004b; Sternad et al., 2010). When practice was examined up to 16 days, invariably noise-tolerant solutions were found first (M. O. Abe & Sternad, 2013; Cohen & Sternad, 2009, 2012). In a similar vein, work on a rhythmic ball bouncing task highlighted that zero-error performance can be achieved in a multitude of ways (Dijkstra et al., 2004; Schaal et al., 1996; Sternad, Duarte, et al., 2000; K. Wei et al., 2007, 2008). Due to its rhythmic nature, there exists a subset of solutions that are dynamically stable, i.e. where small errors or noise are automatically corrected without requiring explicit corrections. These “smart” solutions attenuate the propagation of error and are therefore computationally less demanding. Several studies showed that such dynamically stable solutions require practice and are the hallmark of expert performance (K. Wei et al., 2008). Error scores alone do not reflect this advantageous strategy, and hence, cannot be the only feedback driving performance improvement (Huber, Seitchik, Brown, Sternad, & Harkins, 2015). Therefore, other implicit information or guidance is
needed to “coach” the subject. How can subjects be guided to such dynamically stable solutions?

To date, only relatively few studies in motor neuroscience have used the virtual environment to design interventions that shape behavior in implicit ways, guiding performance to achieve lasting changes in skill. One way to enhance awareness of the desired motion is amplifying the error in the visual feedback. Patton and colleagues have demonstrated the benefit of this method for healthy subjects and also stroke patients, although their task consisted of straight-line reaches, and after-effects remained a critical issue when using the adaptation paradigm (Hasson, Zhang, Abe, & Sternad, n.d.; Milot, Marchal-Crespo, Green, Cramer, & Reinkensmeyer, 2010; Patton, Wei, Scharver, Kenyon, & Scheidt, 2006; Reinkensmeyer & Patton, 2009; Sharp, Huang, & Patton, 2011; Y. Wei, Bajaj, Scheidt, & Patton, 2005). Using a shuffleboard task, Chu et al. (2013) manipulated the variability of the puck releases and showed that behavior can be shaped: decreasing variability by filtering over past trials improved performance, even in severely handicapped children, while increasing variability induced healthy children to perform with more risk awareness. Note that the shuffleboard task was constrained to have no redundancy. Using a redundant line reaching task, Manley, Dayan, and Diedrichsen (2014) attempted to guide subjects to reach in directions for which they receive a monetary reward by adding noise to undesired reaching directions. However, the added noise did not help subjects find the most robust solutions as initially expected. Interestingly, not even a gradient in the noise amplitude guided subjects towards the desired direction, except when subjects were made explicitly aware that noise was added.
It is noteworthy that all of these prior studies on learning and adaptation were focused on discrete movements and none have looked at rhythmic movements. Rhythmic movements are ubiquitous, ranging from locomotion to many forms of tool use. Behavioral, modeling, and neuroimaging results have shown that rhythmic movements follow different principles and may constitute a different “primitive” (Howard et al., 2011; Ikegami et al., 2010; Ronsse et al., 2009; Schaal et al., 2004; Sternad, Dean, & Schaal, 2000; Sternad et al., 2013). Hence, it is likely that also the underlying mechanisms of learning are different. For example, explicit quantitative feedback was not needed to acquire and retain a bimanual rhythmic skill (Park et al., 2013; Park & Sternad, 2015).

This research pursues to develop interventions that accelerate novel skill learning of a rhythmic task in a virtual environment. Going beyond simple reward signals, the study aims to identify elements of performance in a redundant skill to guide learners towards desired dynamically stable and noise-tolerant solutions. The experiment capitalizes on previous research on the task of rhythmically bouncing a ball with a racket that has shown to have redundancy with a subset of solutions displaying dynamic stability (Dijkstra et al., 2004; Ronsse & Sternad, 2010; Schaal et al., 1996; Sternad, Duarte, et al., 2000). In this task, the subject is instructed to rhythmically bounce a virtual ball to a target line using a real racket. Mathematical analysis of a model system indicated dynamically stable solutions, when the racket hit the ball in the decelerating upward swing, assuming sinusoidal motion. This strategy is advantageous because the performer need not adapt his/her racket movements to every small deviation of the ball to maintain successful performance.
Prior experiments by Sternad and colleagues have shown that while novices initially hit the ball with positive racket acceleration, they learn to exploit dynamic stability, as indicated by a shift to negative racket acceleration at impact after approximately 30 minutes of practice (Ehrlenspiel et al., 2010; Huber, Seitchik, et al., 2015; Sternad et al., 2001). Interestingly, performers were not aware of their control of contact, nor of their change in strategy. Dynamic stability has been shown in other rhythmic tasks, the most notable of which is locomotion. This finding is robust across different types of gait, as many animals, including guinea fowl and cockroaches, also exploit dynamic stability in locomotion (Daley, Usherwood, Felix, & Biewener, 2006; Ting, Blickhan, & Full, 1994). In fact, people who are at increased risk of falling, such as amputees or patients with sensory neuropathy, walk slower to improve dynamic stability of the upper body during level walking, even at the cost of increased variability (Beurskens, Wilken, & Dingwell, 2014; Dingwell, Cusumano, Sternad, & Cavanagh, 2000; Dingwell & Marin, 2006).

The purpose of this study was to design a manipulation in the ball bouncing task that guides subjects to these dynamically stable solutions earlier in practice. How can subjects be made aware of these attractor solutions? One approach is to physically shape the subject’s movement during the task using robotic assistance, driving them to exploit the dynamically stable solutions. However, Marchal-Crespo et al. (2014) showed that such haptic guidance actually hampers learning in a similar rhythmic ball bouncing task. Another approach is to manipulate the task itself to guide behavior, without creating obvious dynamic perturbations such as force fields that subjects need to match exactly to be successful. As the ball bouncing task is performed in a virtual environment, the
physical laws that generate the ball movements can be modified to manipulate the attractors in the task. Previously, Morice et al. (2007) showed that shifting the position and velocity of the virtual racket shifted the attractor solutions, but this manipulation created a new perceptual-motor mapping that needed to be matched. Hence, like in visuo-motor adaptation studies, the learned behavior disappeared immediately after removal of the manipulation.

The goal of this research is to guide subjects to better solutions in the present task. The important evaluation of success is that the learned behavior should persist after terminating the intervention. Before the intervention is detailed and the specific hypotheses formulated, the task, the model, and the proposed intervention will be laid out.

5.3 Experimental Methods and Design

5.3.1 The Task and the Model

In the experiment, the subject is instructed to rhythmically bounce a virtual ball to a target line using a real racket. This deceptively simple task requires considerable perceptually-guided coordination to intercept the ball at the right moment and with the right racket velocity to impart the necessary energy to the ball to hit the target height. The core challenge is that control of the ball is confined to the extremely short moments of ball-racket collisions. Further, because impacts occur in a repeating fashion, errors in one impact also determine the next contact: a higher ball amplitude leads to a higher ball velocity at the next contact, that will require a smaller racket velocity at the next contact to compensate. This model task exemplifies rhythmic interaction with an object as is pervasive in tool use, such as hammering, sawing, sweeping, and typing on keyboard.
The model for this task is a well-studied nonlinear dynamical system, originally developed for a particle bouncing on a vibrating surface and then used for a series of human studies (Guckenheimer & Holmes, 1983; Schaal et al., 1996; Sternad et al., 2001; Tufillaro et al., 1992). This simple model consists of a planar surface moving sinusoidally in the vertical direction to repeatedly impact a ball (Figure 5.1). The vertical position of the virtual ball $x_b$ between the $k$th and the $k+1$th racket-ball impact follows ballistic flight:

$$x_b(t) = x_b(t_k) + v_b^+(t - t_k) - g/2 (t - t_k)^2$$

where $t_k$ is the time of the $k$th ball-racket impact, $v_b^+$ is the velocity of the ball just after impact, and $g$ is the acceleration due to gravity (9.81 m/s$^2$). To determine the ball velocity just after impact $v_b^+$, an instantaneous impact is assumed that has energy loss at the collision quantified by a coefficient of restitution $\alpha$:

$$\alpha (v_b^-(t_k) - v_r^-(t_k)) = -(v_b^+(t_k) - v_r^+(t_k))$$

where $v_b$ and $v_r$ are the racket and ball velocities just before (−) and after (+) impact. Further, the mass of the racket is assumed to be much larger than the mass of the ball, such that the racket velocity does not change during impact:

$$v_r^-(t_k) = v_r^+(t_k) = v_r(t_k).$$

Thus, the ball velocity just after impact is determined by:

$$v_b^+(t_k) = (1 + \alpha) v_r(t_k) - \alpha v_b^-(t_k).$$

By assuming sinusoidal racket motion, the racket and ball system is a continuous dynamical system. A Poincare section at the moment of the collision rendered a discrete map with two state variables, ball velocity just after impact $v_b^+$ and racket phase at impact $\theta_k$. 
Local linear stability analysis of this discrete map identified a fixed-point attractor, when racket acceleration at impact $a_r$ satisfied the inequality (Dijkstra et al., 2004; Schaal et al., 1996):

$$-2g \frac{(1 + \alpha^2)}{(1 + \alpha)^2} < a_r < 0$$

For $\alpha = 0.6$ and $g = 9.81$ m/s$^2$, as in the experiment, the range with dynamic stability was between 0 and $-10$ m/s$^2$. Simulations of the ball bouncing map illustrate that when the impact occurs during negative racket acceleration of the upward racket swing (Figure 5.2A), the ball exhibits stable period-1 behavior. The map possesses other attractors besides the period-1 attractor, including “sticking” behavior, where the ball sticks to the racket and follows the racket trajectory. This “sticking” behavior results when the ball-racket impact occurs during positive racket acceleration (Figure 5.2A). Further, non-local Lyapunov stability analyses narrowed the range of acceleration values to $-2$ m/s$^2$ to $-5$ m/s$^2$ for the given parameters $\alpha$ and $g$ (Schaal et al., 1996), although these values are not hard boundaries.

Unlike the ball bouncing map that only describes feedforward dynamics, novice participants who hit with positive racket acceleration are able to compensate for the errors arising from such unstable performance. Based on visual information about the error, they can actively correct for errors by adjusting their racket trajectory to propel the ball either higher or lower than the previous bounce (de Rugy, Wei, Müller, & Sternad, 2003; Siegler, Bardy, & Warren, 2010; K. Wei et al., 2007). However, with practice, participants learn to hit the ball with negative acceleration. By exploiting this efficient solution, small errors need not be corrected, reducing the necessity for active correction (K. Wei et al., 2008).
In the virtual environment, the model was rendered exactly and there were no uncontrolled aspects or simplifying assumptions, such as drag or spin, as would occur in a real experiment.

5.3.2 Design of the Intervention

Given this known dynamics, the question is how the system can be tweaked to guide subjects to more stable behavior? The two complementary principles are penalizing undesirable solutions or enhancing desirable solutions. As Figure 5.2A showed, contact points that ensure stability were on the upper segment of the upward sinusoidal trajectory. To be exact, the period-1 attractor arose from hitting at a segment with racket acceleration between −0 and −10m/s² (Figure 5.3A).

To encourage novices to hit with more negative racket acceleration, the period-1 attractor was “shifted” to a segment of the racket trajectory with more negative acceleration (Figure 5.3B). This manipulation of the dynamically stable solutions was achieved by using the racket velocity 50ms prior impact, as opposed to the racket velocity at impact, to determine the release velocity of the ball. The temporal shift not only made the negative racket acceleration regions more dynamically stable, but also made the positive racket accelerations more unstable, leading to sticking solutions. For the implementation, the racket velocity at impact \( v_r \) was set equal to the racket velocity 50ms before the time of impact \( t_k \). Hence, the ball velocity just after impact was determined by:

\[
v_b^+(t_k) = (1 + \alpha)v_r(t_k - 0.05) - \alpha v_b^-(t_k).
\]

As in the original map, the ball exhibited “sticking” behavior, if the impact occurred during the positive racket accelerations (Figure 5.2B). However, there was an
additional segment of 50ms that produced sticking solutions, where subjects were “penalized”. Only the more negative racket accelerations continued to produce stable period-1 behavior. Given prior findings that humans seek dynamically stable solutions, we expected that subjects performing the task with this manipulation would hit with more negative racket acceleration compared to subjects performing the task under the normal condition.

To test that this state-determined intervention did not only add noise to the ball release, at least in the perception of the subject, but did contain the intended directional information, an additional control condition was included. In this condition, random noise was added to the racket velocity before calculating the ball trajectory. At each bounce, a value was drawn from a Gaussian distribution with mean zero. For better comparison with the time-shifted condition, the standard deviations of the Gaussian were matched with the ones obtained from the velocity shifts in the experimental condition: $\sigma=0.4$ m/s.

Based on these manipulations, the present study tested the following three hypotheses: (1) Subjects who practice with the time-shifted racket velocity learn to hit with negative racket acceleration earlier in practice, compared to those who practiced with no manipulation. (2) After removing the manipulation, they have a higher degree of dynamic stability than those who practice under normal or noise conditions. (3) The attained dynamic stability persists throughout extended practice. Two experiments were conducted to test these hypotheses. In the first experiment, the experimental group practiced for 6 blocks of 4 trials with the manipulation followed by one test block without the manipulation. The purpose of this experiment was to assess the effect of the time-shifted racket velocity on learning and if performance changed upon the removal of the
time-shift manipulation. In the second experiment, the experimental group practiced for 4 blocks followed by 3 test blocks without the manipulation. This experiment further assessed how long this enhanced performance persisted after removing the manipulation.

5.3.3 Participants

A total of 29 students (15 males and 14 females, mean age 20.45 ±3.31 years) from Northeastern University participated in the experiment after signing the consent form approved by the Institutional Review Board of Northeastern University. They received partial fulfillment of a course requirement in exchange for their participation. None had any prior experience with the virtual ball bouncing task. They all self-reported to be right-hand dominant and performed the task with their dominant hand.

5.3.4 Experimental Task and Apparatus

The participant manipulated a real table tennis racket to rhythmically bounce a virtual ball to a target height in a 2D virtual environment (Figure 5.4A). The participant stood 2m in front of a rear projection screen holding the racket in his or her dominant hand. A light rigid rod with two hinge joints was attached to the racket surface, which could translate in the vertical direction while free to tilt around all three axes. The latter was included to minimize friction. However, only the vertical component of the racket displacement moved the virtual racket. As the virtual ball movement was confined to the vertical dimension, the racket movements did not deviate very much in other directions. To measure vertical racket displacement, the rigid rod moved a wheel, whose rotations were registered by an optical encoder at a sampling rate of ~500Hz with a spatial resolution of 0.27mm (Bourns Inc, Riverside, CA). In addition, a wireless accelerometer attached to the center of the racket surface measured the racket accelerations directly at a sampling rate of ~250Hz (Myon 320, Schwarzenberg, Switzerland). A PC (2.4-GHz
Pentium CPU, Windows XP) controlled the experiment and generated the visual stimuli with a graphics card (Radeon 9700, AMD, Sunnyvale, CA). The same PC also acquired the data using a 16-bit A/D card (NI-USB6229BNC, National Instruments, Austin, TX). The delay between real and virtual racket movement was measured in a separate experiment and was on average 22 ± 0.5ms. The images were displayed by a rear-projector (DepthQ-WXGA, Lightspeed Design, Bellevue, WA) consisting of 1024 × 768 pixels with a 60Hz refresh rate.

The vertical ball position, and consequently the maximum ball height of each bounce, was fully determined by the ball velocity, racket velocity, and racket position at impact. In the task model, the coefficient of restitution $\alpha$ was set to 0.6. Just before the virtual ball hit the virtual racket, a trigger signal was sent out to a mechanical brake that acted on the rod. The brake was controlled by a solenoid and applied a brief braking force pulse to the rod to create the feeling of a real ball hitting the racket (Magnet-Schultz type R 16 ×16 DC pull, subtype S-07447). The trigger signal was sent 15ms before the ball–racket contact to overcome the mechanical and electronic delay of the solenoid and brake. The duration of the force pulse (30ms) was consistent with the average impact duration observed in a real ball–racket experiment (Katsumata et al., 2003).

Figure 5.4 shows the virtual environment displayed on the projection screen, which consisted of a virtual racket (horizontal red line, 0.2m x 0.02m), target line (horizontal yellow line, 1.0m x 0.02m), and ball (white circle, 0.02m radius). The vertical position of the virtual racket was determined by the measured position of the real racket; the target line was positioned 1.0m above the minimum racket position. At the start of each trial, the ball was positioned at the left edge of the screen atop the target line and proceeded to
roll along the target line towards the center of the screen. Once it reached the edge of the target line, the ball dropped towards the racket. The participant was instructed to bounce the ball for the duration of the 40s trial such that the maximum ball height of each bounce was coincident with target line. The trajectory of the virtual ball after the ball-racket impact was determined using the ballistic flight and instantaneous impact equations above.

5.3.5 Design and Procedure

In both experiments, all subjects performed a total of 7 blocks, with 4 trials per block, leading to a total of 28 trials. As the duration of each trial was 40s, with a brief break between blocks, the entire experiment lasted approximately 35 minutes.

In Experiment 1, 9 subjects practiced the task with the manipulated racket velocity (time-shifted group), and 9 subjects practiced with no manipulation (control group) (Figure 5.4B). An additional group of 3 subjects practiced with a random noise term added to the racket velocity at impact (noise group). The noise group served as supplementary control to determine that it was the time-dependent nature of the manipulation that presented implicit guidance to dynamic stability, not noise alone. Only 3 subjects were collected as it became immediately evident that they did not understand the result of their actions, and hence, did not change performance. As this random condition was extremely frustrating to subjects, we stopped after collecting 3 subjects. After the 6 practice blocks, all subjects performed one test block of 4 trials with no manipulation to the racket velocity. Subjects were not informed that the conditions changed between the practice blocks and test block.

In order to further assess how much practice with the manipulation was needed to establish dynamic stability and how long this persisted, a second experiment was
conducted. In Experiment 2, 9 subjects practiced with manipulated racket velocity (time-shifted group) for 4 practice blocks, followed by 3 test blocks (Figure 5.4C). The same control group from the first experiment was used for comparison.

5.3.6 Dependent Measures

Error was defined as the signed distance between the maximum ball amplitude and the target line (Figure 5.5). Absolute error was defined as the absolute value of error. The median of absolute error and the interquartile range of the signed error over all bounces in each 40s trial served as performance measures. Shapiro-Wilk tests revealed that on average, the distribution of error and absolute error in each block deviated from normal distribution in approximately 14.5 and 26 out of 28 blocks, respectively, for each subject (Shapiro & Wilk, 1965). Thus, the median and interquartile ranges of these measures were used.

As introduced above, racket acceleration at impact served as the measure of dynamic stability. The racket acceleration signal was measured by an accelerometer on top of the real racket. The signal was then resampled at a fixed frequency of 500 Hz and filtered by a fourth-order Savitzky-Golay filter with a window size of 10ms on both sides (Savitzky & Golay, 1964). Racket acceleration at impact was defined as the racket acceleration in the y-direction 6ms prior to ball-racket impact to avoid capturing any artifacts due to the activation of the mechanical brake. Given that impacts occurred in the upward movement, this temporal interval was conservative, as it biased the estimated value towards positive values. Again, median racket acceleration at impact over all bounces in each 40 s trial was used as Shapiro-Wilk tests revealed that the distribution of racket accelerations in each block was not normal in approximately 9.7 out of 12 blocks for each subject.
The dependent measures of each trial were averaged across trials and blocks for each experimental group.

5.3.7 Statistical Analyses

5.3.7.1 Experiment 1

To assess the effect of the time-shifted racket velocity on learning a 2 (group: control vs. time-shifted) x 6 (block: practice blocks 1 through 6) ANOVA was conducted on each dependent measure (Hypothesis 1). To assess performance changes from the last practice block to the test block, a two-way 2 (group: control vs. time-shifted) x 2 (block: practice block 6 vs. test block) ANOVA was conducted on each dependent measure (Hypothesis 2). Only the time-shifted and control groups were considered in the statistical analyses. The noise group was omitted, as it had only a very limited number of participants.

5.3.7.2 Experiment 2

As in Experiment 1, a 2 (group: control vs. time-shifted) x 4 (block: practice blocks 1 through 4) repeated-measures ANOVA was conducted on each dependent measure (Hypothesis 1). A 2 (group: control vs. time-shifted) x 2 (block: practice block 4 vs. test block 1) ANOVA was conducted on each dependent measure to assess performance changes upon removing the time-shift manipulation (Hypothesis 2). An additional 2 (group: control vs. time-shifted) x 3 (block: test blocks 1 through 3) ANOVA was conducted on each dependent measure to further assess the persistence of performance after removing the time-shift manipulation (Hypothesis 3).

For all ANOVAs, group was a between-subjects factor and block was a within-subjects factor, and the Greenhouse-Geisser correction factor was applied to the within-subject effects (Greenhouse & Geisser, 1959). The significance level was set to $\alpha = 0.05$. 
A test of simple effects was calculated a significant interaction was present. In Experiment 2, one participant in the time-shifted group had an average median racket acceleration at impact across trials that was more than 4 standard deviations above the mean of the time-shifted group. This participant was excluded from the statistical analyses.

5.4 Results

5.4.1 Experiment 1

5.4.1.1 Exemplary Time Series

Figure 5.6 shows exemplary trials from one subject in each experimental group. Each trial lasted 40s and typically contained between 40-50 bounces. The trials shown in Figure 5.6 are the first trial in practice block 1, the last trial in practice block 6, and the first trial in the test block immediately after the manipulations to racket velocity was removed. The data illustrate that despite the initially perturbed performance in the time-shifted group, subjects achieved consistent performance by block 6. This performance persisted without visible change in the test block. The trials of the noise group illustrate how the added noise could lead to sticking solutions throughout practice. Figure 5.7A-C shows the group means of the dependent variables racket acceleration, absolute error and variability of error, plotted across the 6 practice blocks and the 1 test block for the 3 experimental groups.

5.4.1.2 Learning with Time-Shift Manipulation

Consistent with prior results, subjects of all 3 groups initially hit with positive racket acceleration in block 1 (time-shifted: $M = 2.66\text{m/s}^2$, $SD = 4.87\text{m/s}^2$; control: $M = 2.40\text{m/s}^2$, $SD = 2.06\text{m/s}^2$). In subsequent blocks, all subjects decreased their racket
acceleration, albeit at different rates and to different degrees (Figure 5.7A). Hypothesis 1 stated that subjects who practice with the time-shifted racket velocity learn to hit with negative racket acceleration earlier in practice, compared to those who practiced with no manipulation. The ANOVA on median racket acceleration at impact revealed a significant main effect of block, $F(1.65, 26.42) = 28.69, p < .001$, as well as group, $F(1, 16) = 8.03, p = .012$. These main effects were qualified by a significant interaction between group and block, $F(1.65, 26.42) = 3.65, p = .047$. A test of simple effects revealed that the time-shifted group had significantly lower racket accelerations at impact compared to the control group in blocks 3 through 6 ($ps < .015$). These results indicate that the time-shifted manipulation could guide subjects to hit with negative racket acceleration earlier in practice (Hypothesis 1). While the noise group was not subjected to statistical testing, the median racket acceleration at impact remained positive over practice, indicating that the time-dependent nature of the manipulation was necessary for the desired change in behavior (Figure 5.7A).

To determine if practicing with the time-shifted racket velocities interfered with task performance, ANOVAs on error measures were conducted. The ANOVA on median absolute error revealed a main effect of block, $F(1.68, 26.88) = 16.2, p < 0.001$. As expected, median absolute error decreased with practice for both groups (Figure 5.7B). However, neither the main effect of group, $F(1, 16) = 1.58, p = .23$, nor the Group x Block interaction, $F(1.68, 26.88) = 1.45, p = .25$, were significant. The statistical results of the mean absolute error were mirrored in the analysis of the interquartile range of error, as seen in Figure 5.7C. The ANOVA on interquartile range of error revealed a significant main effect of block, $F(3.92, 52.61) = 11.11, p < 0.001$, but not for group, $F(1,
The Group x Block interaction was also non-significant, $F(3.92, 52.61) = .67, p = .59$. In short, while both groups improved overall task performance over time, there were no significant differences between the time-shifted and control groups in these error measures. Visual inspection of Figures 7.7B-C shows that the noise group has performed visibly worse over the course of practice compared to the control and time-shifted groups.

5.4.1.3 Removing Time-Shift Manipulation

To determine whether performance changed upon removal of the time-shifted manipulation, the dependent measures in practice block 6 and the test block were compared (Figure 5.7A-C). The ANOVA on median racket acceleration at impact revealed a significant main effect of group, $F(1, 16) = 18.44, p = .001$, but not for block, $F(1, 16) = 1.17, p = .30$. The analysis also revealed a significant interaction between group and block, $F(1, 16) = 7.91, p = 0.013$. A test of simple effects revealed that the time-shifted group significantly increased racket acceleration at impact from practice block 6 ($M = -4.61\text{m/s}^2, SD = 1.1\text{m/s}^2$) to the test block ($M = -3.73\text{m/s}^2, SD = 1.13\text{m/s}^2$), $p = .014$, whereas the control group did not, $p = .24$. While the significant increase in racket acceleration in the timed-shifted group was inconsistent with Hypothesis 2, it should be kept in mind that, while statistically different, the acceleration values did not differ much in their physical meaning. As stated above, a wide range of negative values ensures dynamic stability and an “optimal region” as determined by Lyapunov stability analysis spans $-2$ to $-5\text{m/s}^2$ (Schaal et al., 1996). While the time-shifted group hit with racket acceleration values in this optimal range during the test block ($M = -3.73\text{m/s}^2, SD = 1.13\text{m/s}^2$), the control group did not ($M = -1.20\text{m/s}^2, SD = 2.06\text{m/s}^2$).
The ANOVA on median absolute error did not reveal any significant main effects (block: \( F(1,16) = 1.39, p = .26 \); group: \( F(1,16) = .013, p = .91 \)), nor a significant interaction, \( F(1,16) = 2.09, p = .17 \). Similarly, the ANOVA on the interquartile range of error did not reveal any significant main effects (block: \( F(1,16) = .53, p = .48 \); group: \( F(1,16) = .007, p = .94 \)), nor a significant interaction, \( F(1,16) = 1.52, p = .24 \). Thus, removing the manipulation did not interfere with task performance.

### 5.4.2 Experiment 2

Figure 5.8A-C shows the group means of dependent measures plotted across the 4 practice blocks and the 3 test blocks. The control group was the same group from Experiment 1.

#### 5.4.2.1 Learning with Time-Shift Manipulation

Just as in Experiment 1, the time-shifted group initially hit with positive racket acceleration in block 1 (\( M = 2.20\, \text{m/s}^2, SD = 3.81\, \text{m/s}^2 \)). The ANOVA on median racket acceleration at impact in the 4 practice blocks revealed a significant main effect of block, \( F(1.92, 28.76) = 16.07, p < .001 \) (Figure 5.8A). The main effect of group did not reach significance, \( F(1, 15) = 3.63, p = .076 \). These effects were qualified by a significant interaction between Group and Block, \( F(1.92, 28.76) = 3.44, p = .047 \). A test of simple effects revealed that the time-shifted group had significantly lower racket acceleration at impact compared to the control group in blocks 2 and 4 (\( ps < .015 \)). This result indicates that even after just 4 practice blocks, the time-shifted manipulation guided subjects to hit with negative racket acceleration earlier in practice (Hypothesis 1).

The ANOVA on median absolute error revealed a main effect of block, \( F(1.91, 28.62) = 18.05, p < .001 \) (Figure 5.8B). However, neither the main effect of group, \( F(1, 15) = 2.54, p = .13 \), nor the Group x Block interaction, \( F(1.91, 28.62) = 1.63, p = .21 \),
were significant. The ANOVA on interquartile range of error similarly revealed a significant main effect of block, $F(2.46, 36.91) = 3.04, p = .050$, but not for group, $F(1, 15) = .32, p = .58$ (Figure 5.8C). The Group x Block interaction also was not significant, $F(2.46, 36.91) = 1.17, p = .33$. As in Experiment 1, there were no significant differences between the time-shifted and control groups in these error measures over the practice blocks.

### 5.4.2.2 Removing Time-Shift Manipulation

To determine whether performance changed upon removal of the time-shifted manipulation, the dependent measures in practice block 4 and the test block 1 were compared (Figure 5.8A-C). The ANOVA on median racket acceleration at impact revealed a significant main effect of group, $F(1, 15) = 7.94, p = .013$. Pairwise comparisons revealed that in the time-shifted group hit with significantly lower racket acceleration than the control group in both practice block 4 and test block 1 ($p_s < .04$) (Hypothesis 2). The main effect of block was marginally significant, $F(1, 15) = 4.47, p = .052$, and the interaction between group and block was not significant, $F(1, 15) = .001, p = .98$.

The ANOVA on median absolute error revealed a significant main effect for block $F(1, 15) = 4.60, p = .049$, but not for group, $F(1, 15) = 1.43, p = .25$. The interaction was also not significant $F(1, 15) = .33, p = .57$. Similarly, the ANOVA on interquartile range of error revealed a significant main effect for block, $F(1, 15) = 5.93 p = .028$, but not for group, $F(1, 15) = .36, p = .56$. The interaction was also not significant, $F(1, 15) = .64, p = .44$. Consistent with Experiment 1, these results indicate that removing the manipulation did not interfere with task performance.
5.4.2.3 Persistence

In Experiment 2, the time-shifted group performed 3 test blocks under normal task conditions after removing the perturbation. The ANOVA on median racket acceleration at impact in the 3 test blocks revealed a significant main effect of group, $F(1, 15) = 7.71$, $p < .014$, as seen in Figure 5.8A (Hypothesis 3). Pairwise comparisons revealed that the time-shifted group hit with significantly lower racket acceleration than the control group in test blocks 1 and 2 ($ps < .035$). Neither the main effect for block, $F(1.29, 19.39) = 1.16$, $p = .31$, nor the interaction between group and block were significant, $F(1.29, 19.39) = 3.68, p = .061$. As previously reported, by the end of the experiment, the control group hit with racket acceleration values that were still outside the optimal region of dynamic stability. Like in Experiment 1 and consistent with Hypothesis 3, the time-shifted group in Experiment 2 learned and maintained hitting with racket acceleration at impact in this region until the end of the experiment ($M = -2.61m/s^2$, $SD = 1.55m/s^2$ in test block 3).

The ANOVAs on median absolute error and interquartile range of error did not reveal significant main effects for block and group, nor interactions in the test blocks, $Fs < 1, ps > .50$.

5.5 Discussion

This study examined how the manipulations of a task in a virtual environment can implicitly guide subjects towards a desired behavior that also persisted after removal of the guidance. Using the rhythmic perceptual-motor skill of bouncing a ball, this experiment introduced a task-based intervention to subtly shape performance. Our approach capitalized on virtual environments, which has become a prominent tool in
motor neuroscience and rehabilitation to deliver performance feedback that can go beyond delivering quantitative information after completion of performance (Holden & Todorov, 2002; Huber et al., 2010; Lange et al., 2012). The results supported our three hypotheses: Subjects could be guided to learn the dynamically stable solution faster than controls (*Hypothesis 1*). Immediately after the manipulation was removed, they maintained their performance under normal conditions, without any undesired after-effects (*Hypothesis 2*). This enhanced performance persisted over extended practice (*Hypothesis 3*).

Unlike most previous studies on learning and adaptation, we chose a rhythmic task, complementing the numerous learning and adaptation studies on discrete reaching movements. Based on previous theorizing and experimental support, it may be expected that learning rhythmic movements has different characteristics and may obey different principles (Hogan & Sternad, 2007, 2013). This study focused on dynamic stability, a characteristic inherent to rhythmic movements, ranging from ball bouncing to locomotion. Previous theoretical and experimental research on the ball bouncing task showed that dynamic stability afforded a solution that obviated error correction and was less sensitive to noise. Skilled experts robustly converged to these solutions, characterized by negative racket accelerations, as shown in several different experimental set-ups and instructional conditions (Huber, Seitchik, et al., 2015; Katsumata et al., 2003; Schaal et al., 1996; K. Wei et al., 2007, 2008). Importantly, subjects are not aware of how their strategy changes with practice. Hence, explicit information about the desired strategy or the central variable is probably less effective or even unnecessary, although this remains to be tested.
This study developed a subtle intervention that could steer subjects to the dynamically stable behavior. Unlike in many motor adaptation studies using force fields or visuo-motor rotations, where the after-effects are very short-lived when the manipulation was removed, the learned behavior in this study persisted. In Experiment 1, the test block showed that racket accelerations stayed negative, although there was a significant, yet small, increase from \(-5.32 \text{m/s}^2\) to \(-4.40 \text{m/s}^2\). However, this increase should not be over interpreted, as both values remained within the range of stability between \(-2\) and \(-5\text{m/s}^2\) (Schaal et al. 1996). In fact, prior studies showed that experienced subjects converged to hitting with a racket acceleration value of approximately \(-3 \text{m/s}^2\) (K. Wei et al., 2008). Hence, this increase may be interpreted in line with this previous observation. Experiment 2 further tested the persistence by including three test blocks that showed again that the impacts maintained the signature of dynamic stability. It should be pointed out that 3 blocks with 12 trials of approximately 60 bounces each amounted to \(~720\) single bounces and presents a relatively long test phase. By the end of the experiment, the two time-shifted groups learned to use solutions with local linear stability, whereas the control group only learned solutions outside of this desired range. While the results of these two experiments demonstrate that this efficient behavior persisted after removing the manipulation, further investigation is needed to determine if it is retained in the longer term, and when there are breaks between the practice and the test blocks.

5.5.1 Going Beyond Error to Measure Task Performance

It should be highlighted that over the course of practice, there was no significant difference in the error measures between the control group and the time-shifted group. Hence, analysis of error alone would not have differentiated between the two groups, as
these performance changes differed from the paradigmatic error-based learning (Diedrichsen, White, Newman, & Lally, 2010). As introduced above, bouncing a ball rhythmically has redundancy and low error can be achieved with different strategies. While the dynamically stable strategy is advantageous as it is more noise-tolerant, it is not necessary for task achievement. Previous work highlighted how novices and experts employ different degrees of error correction versus exploitation on these passive error compensation mechanisms (K. Wei et al., 2007, 2008). It should also be pointed out that hitting with more negative racket acceleration was not an immediate response to the manipulation. Rather, it took several blocks of practice until the time-shifted group had reached significantly lower negative racket accelerations at impact.

The observation that the errors did not differ between the experimental and the control group raised the question: what signaled and guided subjects to the dynamically stable behavior? We previously argued that the stable solutions are computationally less demanding because, in principle, they do not require active monitoring of the ball trajectory to correct errors. In contrast, under unstable conditions, continuous corrections of perceived errors in the ball amplitude of the previous bounce are required. Such continuous monitoring requires perceptual and control processes, which subjects might sense as computational effort and seek to gradually minimize. Such arguments are consistent with studies that have shown support for effort minimization (e.g. Emken et al. 2007).

Another type of effort that may play a role in optimizing performance is mechanical energy. Considering one bounce in isolation, the maximum amplitude of the ball is determined by racket velocity at contact, together with the pre-impact ball velocity.
and the absolute height of the impact. Hence, it may be argued that subjects should hit the ball at peak velocity of the racket trajectory to impart maximum energy to the ball for a given racket amplitude. Subjects may then decrease the racket amplitude, while still impacting the ball at sufficiently high peak velocities. This was actually proposed in previous work on juggling robots, but has never been observed in our human experiments (Bühler, Koditschek, & Kindlmann, 1990, 1994). Nevertheless, to examine whether racket amplitude decreased with practice, facilitated by the time-shift condition, racket amplitudes were analyzed. Consistent with prior findings, the racket amplitudes neither increased nor decreased throughout practice. This permitted the conclusion that mechanical efficiency did not play a role.

5.5.2 Time-Shifted Manipulation is Different than Adding Random Noise

The random noise condition was added to test whether it was indeed the subtle velocity manipulation that had the desired effect on performance, and it was not only perceived as noise. As all measures show, performance in the noise condition was significantly the worst. In fact, subjects in this group very quickly showed frustration, because none of their attempts resulted in the desired performance. Hence, only three subjects were collected as the prolonged task performance with random outcome was thoroughly annoying to subjects. One other rationale behind adding extrinsic noise was that such amplification of intrinsic neuromotor noise may aid in the exploration of the execution space to discover the more advantageous solutions. However, this rationale proved not effective in the current experiment. One reason may have been that the added noise was relatively large. After the experiment, subjects in the noise-group reported a sense of helplessness, as they felt that they had no control over their own performance. Manley et al. (2014) similarly reported in a redundant line-reaching task that neither
adding nor amplifying noise led subjects to discover the optimal solution, unless they were explicitly aware of the task-relevant variable.

5.5.3 Designing Interventions for Persistent Behavior

This persistence occurred because subjects most likely did not perceive any change in the task or environment when the manipulation was removed. One reason why subjects did not notice the removal of the manipulation is that the manipulation did not cause a change in the visuo-motor mapping. A previous experiment by Morice et al. (2007) on the same rhythmic task did not achieve persistent behavior after removal of a manipulation that altered the visuo-motor mapping. With the goal to examine how subjects learned new attractor solutions, they temporally shifted the racket position and velocity, which altered the visual display of the racket. This created a mismatch between proprioceptive and visual feedback and a new perceptual-motor mapping had to be learned. Hence, similar to visuo-motor adaptation studies, this acquired behavior disappeared very fast after the display returned to normal conditions. To avoid such fast return to baseline behavior, the present study only manipulated the racket velocity, which did not alter the visual display of the racket and introduce new visuo-motor mappings.

A second reason why the removal of the perturbation did not affect behavior is that the changed behavior also resulted in successful performance of the task under normal conditions. Even though subjects experienced initial difficulties when they performed with their typical novice strategy, mostly sticking solutions (Figure 5.5), they quickly learned to hit the ball later in the decelerating racket trajectory, which led to the desired decrease in the error and its variability. By the time the manipulation was removed, subjects had found solutions that led to dynamic stability both in the manipulated and
unmanipulated versions task. Hence, removing the manipulation did not increase error and variability, and subjects did not perceive a reason to alter their behavior.

5.5.4 Implications for Motor Rehabilitation

While rhythmic ball bouncing is an experimental toy task, it has many similarities to walking. Most centrally, the contact fully determines the flight phase. Dynamic stability has also been a central research topic in the area of passive dynamic walking (Garcia, Chatterjee, Ruina, & Coleman, 2007; McGeer, 1990). A recent series of studies by Reisman and colleagues on walking in post-stroke patients has demonstrated the utility for such implicit interventions on gait asymmetries (Reisman, McLean, Keller, Danks, & Bastian, 2013; Reisman, Wityk, Silver, & Bastian, 2007; Reisman et al., 2009). Using a split-belt treadmill, patients walked on two parallel belts and their asymmetries in step length during locomotion were exaggerated by moving the belt for the paretic leg at a slower pace than the non-paretic leg. Increasing the asymmetries drove the nervous system to make the necessary corrections in step length and patients achieved better symmetry between the two step lengths. Furthermore, this improvement transferred to overground walking, which Reisman et al. (2009) attributed to the fact that the improved symmetry in step length may have been beneficial in terms of energy cost, balance, or efficiency. In contrast, when control subjects completed the same split-belt training, the training induced asymmetries that did not transfer to overground walking. Consistent with the findings of the current study, these results suggest that interventions for motor rehabilitation and motor learning must guide subjects towards the correct behavior in unaugmented task conditions.

Robot-assisted therapy is another example of an intervention that guides patients toward the correct behavior. Unlike robot-assisted therapy for the upper limbs, the robotic
interventions for improving locomotion have fallen short so far. The most prominent robotic intervention for locomotion is to passively move the limbs through a predefined kinematic pattern. However recent studies have shown that this type of training does not improve gait (Hidler et al., 2009; Hornby et al., 2008). Marchal-Crespo et al. (2014) similarly showed that in the rhythmic ball bouncing, such robotic guidance of the racket trajectory actually hampered learning. This guidance contrasts with the time-shifted manipulation, where the stable solution was not imposed on the subjects. The subjects were still actively engaged in the learning process, which has proved to be critical for motor recovery (Lynch et al., 2005).

Furthermore, the manipulation provided minimal intervention as it only affected the critical instant, when the ball impacted the racket, as opposed to the entire racket trajectory. Prior studies on ball bouncing have also demonstrated the importance of ball-racket impact in behavior. For instance, in contrast to haptic guidance throughout the entire racket trajectory, haptic feedback at the ball-racket impact does improve performance and learning (Ankarali, Tutkun Sen, De, Okamura, & Cowan, 2014; Sternad et al., 2001). An intervention that focuses on physical contacts has also been shown a promising direction for robotic assistance in locomotor recovery. Ahn and Hogan (2012) have shown that periodic torque pulses to the ankle during gait can increase walking cadence. Similar to the ball bouncing task, exploiting the natural stability of walking trajectories may largely govern locomotion patterns, as suggested by research using passive-dynamic machines (Collins, Ruina, Tedrake, & Wisse, 2005).
5.6 Conclusions and Outlook

With the time-shifted manipulation, the learned behavior persisted because removing the manipulation did not change the visuo-motor mapping and did not decrease task performance. These design principles should be considered when developing interventions for motor rehabilitation, where persistence and ultimately retention of the learned behaviors is critical. While this study could only make a case in point, it highlights that feedback to the learner can and should take on more than error information about the outcome. In learning real world skills, we are typically guided by parents, teachers, and coaches, and their guidance rarely comes in the form of an error score, or an explicit statement what the optimal solution is. Laymen and even coaches all too frequently do not have the vocabulary, let alone the quantitative measures to provide exact description of the target skill or their deviations. Instead, they provide instruction for how it should feel or look to the performer. In contrast, current research and virtual reality-based rehabilitation practices emphasize error- and reward-based learning, which is effective in many experimentally controlled tasks (M. Abe et al., 2011; Galea et al., 2015; Holden & Todorov, 2002; Huber et al., 2010; Winstein & Schmidt, 1990). However, activities of daily living are almost invariably redundant tasks that require additional and subtler guidance to help the learner or patient identify the mapping between execution and task outcome. It would be desirable to further develop principles of such implicit guidance that go beyond the error-based approach and develop manipulations that can be applied for rehabilitation through virtual environments.
5.7 Figures

Figure 5.1. Model of the racket-ball system. The vertical ball position between each instantaneous impact follows ballistic flight, which depends on three execution variables: ball $v_b^-$ and racket $v_r^-$ velocities just before impact and racket position $x_r$ at impact.
Figure 5.2. (A) Simulation of the ball-racket system. Assuming sinusoidal racket movement, the racket trajectory has a segment with positive acceleration followed by negative acceleration before its peak position during the upward swing. Ball trajectories with initial contact at different phases of the upward swing (from 1.51π rads to 2.49π rads at an interval of .3 rads) were simulated. When the racket impacted the ball during the decelerating portion of the racket’s upward motion, the ball-racket system was dynamically stable. All simulations with initial contact during negative racket accelerations led to the same stable ball amplitude without requiring any changes in the racket trajectory. If the ball impacted the racket during the accelerating portion of the racket’s upward motion, the system was unstable. All simulations with initial contacts during positive racket acceleration led to unstable behavior, where the ball finally stuck to the racket. The only way to achieve and maintain a stable pattern was to correct for errors in the ball amplitude by a change in the racket trajectory. (B) Simulation of the ball-racket system with time-dependent perturbation. If the ball impacted the racket during the accelerating portion of the racket’s upward motion, the system was unstable. When the racket impacted the ball during the decelerating portion of the racket’s upward motion, the system was unstable if the racket phase was less than .05ω; otherwise, the perturbed ball-racket system was dynamically stable.
Figure 5.3. Dynamically stable solutions (gray) depicted on the assumed sinusoidal racket trajectory under the (A) control condition and (B) time-shifted racket velocity condition.
Figure 5.4. (A) Side and front view of the virtual experimental setup for ball bouncing. Participants were positioned in front of a screen and manipulated a real table tennis racket to rhythmically bounce a virtual ball to a target height in a 2D virtual environment. (B) Schematic of Experiment 1 design. Each group performed of 6 practice blocks with respective manipulations to racket velocity and one test block with no manipulation. Each block consisted of 4 trials, and each trial lasted 40 seconds long. (C) Schematic of Experiment 2 design. The time-shifted group performed of 4 practice blocks and 3 test blocks with no manipulation. The control group was the same from Experiment 1.
Figure 5.5. Exemplary time series of racket (black) and ball (gray) to illustrate dependent measures. Error was defined as the unsigned difference between the target height and the maximum ball amplitude at each bounce. Racket acceleration at impact was defined as the racket acceleration 6ms before the ball-racket impact of each bounce.
**Figure 4.** Exemplary trials of three subjects (one per experimental group) in practice blocks 1 and 6 and the test block. The first 25s of the 40s time series of racket (gray) and ball (black) position are depicted. The dashed line indicates the target height.
Figure 5.7. Dependent measures of Experiment 1 over blocks. The measure of each trial were averaged across blocks and finally averaged over the subjects in each experimental group. Each point represents the group average per block, and the error bar represents the standard error across subjects in each group. (A) Median racket acceleration. The dynamically stable solutions between -2m/s² and -5m/s² (shaded region below the dashed line) are desired. (B) Median absolute error. (C) Interquartile range of signed error.
Figure 5.8. Dependent measures of Experiment 2 over blocks. The measure of each trial were averaged across blocks and finally averaged over the subjects in each experimental group. Each point represents the group average per block, and the error bar represents the standard error across subjects in each group. (A) Median racket acceleration. The dynamically stable solutions between -2m/s² and -5m/s² (shaded region below the dashed line) are desired. (B) Median absolute error. (C) Interquartile range of signed error.
6. The Effect of Stereotype Threat on Performance of a Rhythmic Motor Skill

One key difference between virtual and robotic rehabilitation versus traditional therapy is the human factor. The subtle psychological and motivational cues from the therapist are likely to be very important for the success of the therapy. It may even be argued that the ability to intentionally and effectively employ these cues is what ultimately makes a successful therapist. Even when a patient engages in virtual and robotic rehabilitation, the therapist’s instruction and motivational attitude remains an essential element of success (Hall, Ferreira, Maher, Latimer, & Ferreira, 2010; Maclean & Pound, 2000). In this study, the effect of increased motivation through a verbal instruction on skill performance is assessed. While there are many different ways to increase or decrease motivation, this study employed a stereotype threat instruction to manipulate motivation. The study used a rhythmic ball bouncing task that has proven an effective test bed in previous research and made theoretically-grounded kinematic measures available.

I contributed to the experimental design, data analysis, statistical analysis, and interpretation of results.

6.1 Abstract

Many studies using cognitive tasks have found that stereotype threat, or concern about confirming a negative stereotype about one’s group, debilitates performance. The few studies that documented similar effects on sensorimotor performance have used only relatively coarse measures to quantify performance. Three experiments tested the effect of stereotype threat on a rhythmic ball bouncing task, both at the novice and skilled level.
Previous analysis of the task dynamics afforded more detailed quantification of the effect of threat on motor control. In this task, novices hit the ball with positive racket acceleration, indicative of unstable performance. With practice, they learn to stabilize error by changing their ball-racket impact from positive to negative acceleration. Results showed that for novices, stereotype threat potentiated hitting the ball with positive racket acceleration, leading to poorer performance of stigmatized females. However, when the threat manipulation was delivered after having acquired some skill, reflected by negative racket acceleration, the stigmatized females performed better. These findings are consistent with the mere effort account that argues that stereotype threat potentiates the most likely response on the given task. The study also demonstrates the value of identifying the control mechanisms through which stereotype threat has its effects on outcome measures.

6.2 Introduction

Stereotype threat refers to the concern that is experienced when one feels "at risk of confirming, as self-characteristic, a negative stereotype about one's group" (Steele & Aronson, 1995). A large number of studies have found that concern about confirming the relevant negative stereotype negatively impacts the performance of the stigmatized individuals (Aronson, Fried, & Good, 2002; Ben-Zeev, Fein, & Inzlicht, 2005; Blascovich, Spencer, Quinn, & Steele, 2001; Brown & Pinel, 2003; Davies, Spencer, Quinn, & Gerhardstein, 2002; Jamieson & Harkins, 2007; Johns, Schmader, & Martens, 2005; Schmader & Johns, 2003; S. J. Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995). A variety of cognitive tasks have been used to study these debilitating effects, including GRE verbal and quantitative problems (Steele & Aronson, 1995), tests of
memory (Hess, Auman, Colcombe, & Rahhal, 2003), GMAT problems (Quinn & Spencer, 2001), mental rotation problems (Martens, Johns, Greenberg, & Schimel, 2006), the Stroop Color-Word task (Jamieson & Harkins, 2011), and reading span tasks (Mazerolle, Régner, Morisset, Rigalleau, & Huguet, 2012).

In contrast, very few studies have examined the effect of stereotype threat on the performance of sensorimotor tasks. Three studies have examined golf putting (Beilock, Jellison, Rydell, McConnell, & Carr, 2006; Stone, Lynch, Sjomeling, & Darley, 1999; Stone & McWhinnie, 2008), one investigated soccer dribbling (Chalabaev, Sarrazin, Stone, & Cury, 2008), one driving in a simulator (Yeung & von Hippel, 2008), and one basketball free throw shooting and tennis serving (Hively & El-Alayli, 2014). Not only have few tasks been studied, but also performance on these tasks has been quantified with only relatively coarse measures. For example in the soccer dribbling task, Chalabaev et al., (2008) evaluated overall speed and found that females under stereotype threat completed the drill significantly more slowly than those without threat. While this finding supports the argument that the stereotype has a negative effect on sensorimotor performance, these results are mute about how stereotype threat affects motor performance. In soccer dribbling, slower performance can result either from being slow and cautious, but making minimal errors, or from moving fast, but making hasty mistakes that then need correcting. In golf putting, stigmatized individuals have been shown to be less accurate (Beilock et al., 2006) and require more strokes (Stone & McWhinnie, 2008) to putt a ball to a target hole. As in soccer, these outcome measures say little about how stereotype threat affects sensorimotor control: the increased number of strokes does not reveal whether more cautious behavior or larger errors led to specific changes in the golf
swing without additional measures of sensorimotor control. Finally, a recent study by Hively and El-Alayli (2014) investigated the effect of stereotype threat on collegiate athletes in two different sensorimotor tasks, shooting a basketball and serving a tennis ball, by combining their performances with z-scores.

Performance on complex sensorimotor tasks is the result of a variety of motor, as well as cognitive, processes, and global performance measures alone cannot identify how stereotype threat may affect them. In fact, even if stereotype threat has little or no visible effect on the primary outcome measure, it may still affect the underlying control processes. For example, Chalabaev, Sarrazin, Fontayne, Boiché, and Clément-Guillotin, (2013) found that, on a ballistic contraction task, ST did not affect the primary performance measure, maximum force production, but did influence the peak rate of force production, suggesting some effect on control processes. To investigate how stereotype threat affects sensorimotor control, it is necessary to derive measures that more directly reflect these control processes. For ball tasks such as golf putting, soccer dribbling, tennis serving, and basketball shooting overall performance is determined by the motor actions of the human as well as the dynamics of the ball. Extracting measures that reflect control requires an accurate model of the ball and the external environment.

6.2.1 Rhythmic Ball Bouncing Task

In the present work, we used an experimental task developed by Sternad and colleagues where the participants rhythmically bounced a ball with a racket (Schaal et al., 1996; Sternad et al., 2014; Sternad, 1999, 2006). Participants manipulate a real table tennis racket to bounce a virtual ball to a target height, with movements restricted to the vertical direction (Figure 6.1). The deviation of the maximum ball height from the target height for each bounce served as the outcome measure or error that participants are
instructed to minimize. Because the ball motion was simulated in a virtual environment, there were no uncontrolled aspects as would occur in a real-life version of the task; the ball dynamics were completely known. Knowing the exact physical model of the task allowed us to obtain underlying measures critical for control in addition to the typical performance measure, error.

To extract variables that delineate potential control strategies, a mathematical model of this task was developed and subsequently analyzed (Dijkstra et al., 2004; Sternad, Duarte, et al., 2000; K. Wei et al., 2007). The model consisted of a planar surface performing periodic vertical movements repeatedly impacting a ball (see Appendix A). The model analysis of this nonlinear dynamical system showed that the system has dynamically stable solutions, meaning that small errors or perturbations die out by themselves. This approach is advantageous because the performer need not adapt his/her racket movements to every small deviation of the ball to maintain successful performance. Importantly though, dynamic stability is only achieved when the racket hits the ball during the decelerating portion of the racket’s upward motion. Figure 6.2 illustrates this behavior for simple sinusoidal movement of the racket. Note that for this argument the racket movement does not need to be strictly sinusoidal, only periodic, as only the segment of the ball-racket contact matters. Hence, hitting the ball with negative acceleration is an efficient solution to this task, as small errors need not be corrected.

To make this property of the task more intuitive, consider the following: while hitting the ball with negative racket acceleration, balls that hit the racket earlier (such as after a lower ball amplitude in the preceding bounce) are hit with relatively higher velocity. This in turn leads to a higher ball amplitude on the following bounce, which is
equivalent to an automatic correction of the previous low amplitude bounce. Conversely, balls that hit the racket later (such as after an amplitude overshoot) are hit with relatively lower velocity, which leads to a lower ball amplitude on the next bounce.

Previous research has shown that participants learn to exploit dynamic stability (Dijkstra et al., 2004; Huber & Sternad, 2015; Sternad et al., 2001; Sternad, 2006; K. Wei et al., 2008). Novices initially hit the ball during the accelerating portion of the racket’s upward motion to impart energy to the ball in the upward direction. In contrast to the dynamic stability achieved with ball-racket impacts during the decelerating racket trajectory, hitting the ball with an accelerating trajectory produces unstable performance and errors amplify from one bounce to the next (Figure 6.2B-C). To compensate for such unstable performance, the novice participants can actively correct for errors by adjusting their racket trajectory to propel the ball either higher or lower than the previous bounce, based on visual information about the error (de Rugy et al., 2003; Siegler, Bazile, & Warren, 2013; K. Wei et al., 2007). However, with practice, participants learn to hit the ball with negative acceleration, reducing the necessity for active correction of errors (K. Wei et al., 2008). Thus, this analysis identifies a control variable, racket acceleration, through which stereotype threat may have its effect on the outcome variable, error.

6.2.2 Theories on Stereotype Threat

The effect of stereotype threat on task performance has been conceptualized in two lines of work, one emphasizing performance on cognitive tasks (Schmader, Johns, & Forbes, 2008) and the other performance on sensorimotor tasks (Beilock et al., 2006). (Schmader et al., 2008) have proposed a model that is primarily aimed at accounting for debilitation on cognitive tasks. This model incorporates cognition, affect, and motivation and how they impact performance through their effects on working memory efficiency.
Performance is debilitated because concern about fulfilling the stereotype occupies working memory resources that could be used for task performance.

This model also incorporates a separate pathway for performance on sensorimotor tasks, thereby including results by Beilock and her colleagues (Beilock et al., 2006). Beilock argues that stereotype threat and performance pressure result in a two-pronged effect whereby people not only worry about the situation, thereby depleting working memory as in (Schmader et al., 2008), but also explicitly monitor their performance in order to ensure optimal performance (DeCaro, Thomas, Albert, & Beilock, 2011). Beilock has focused on the performance of skilled athletes, and has found that pressure leads to the monitoring of performance, which disrupts well-learned, proceduralized behavioral sequences (Beilock & Carr, 2001; Beilock et al., 2006).

Schmader et al. (2008)’s account incorporates motivation, but only indirectly, as it suggests that its effects manifest themselves by impacting efficiency of working memory. Beilock’s account (Beilock et al., 2006) argues that threatened participants are motivated to minimize mistakes, but the cause of the performance debilitation is the resulting step-by-step focus on task execution. In the current work, we focus on the effect of stereotype threat on motivation and its direct effects on control and task performance by adopting Jamieson and Harkins (2007)’s mere effort account.

6.2.3 Mere Effort: A Motivational Account

The mere effort account was suggested by Harkins (2006)’s analysis of the effect of the potential for evaluation on performance. This account argues that the potential for evaluation motivates participants to want to perform well, which potentiates whatever response is prepotent, or most likely to be produced, on the given task. For example, on the Stroop Color-Word Task (Stroop, 1935), the prepotent response is to state the color
word, as opposed to the correct response, naming the color in which the word is printed. If the prepotent response is incorrect and participants do not know, or lack the knowledge or time required for correction, performance is debilitated (Figure 6.3). However, if the prepotent response is correct, or if participants are able to recognize that their prepotent tendencies are incorrect and are given the opportunity to correct, performance will be facilitated. Harkins and his colleagues have found support for these predictions on the Remote Associates Task (Harkins, 2006), anagrams (McFall, Jamieson, & Harkins, 2009) (Experiment 1), the Stroop task (McFall et al., 2009, Experiments 2 and 3) and the antisaccade task (McFall et al., 2009, Experiment 4).

For example, Witte and Freund (2001) found that when solving anagrams, the initial tendency was to try consonants in the first position. (McFall et al., 2009, Experiment 1) argued that subjecting participants to the potential for evaluation should potentiate this prepotent response. As a result, participants subject to evaluation should be better at solving anagrams of words that begin with consonants, but worse at solving anagrams that begin with vowels than participants not subject to evaluation. This is what they found. The same process (potentiation of the prepotent response) led to better performance in one case (words that began with consonants), but worse performance in the other (words that began with vowels).

Jamieson and Harkins (2007) argued that stereotype threat, like the potential for evaluation, motivates threatened participants to want to perform well. In this case, the motivation to counter the negative stereotype brings into play the same processes that are implicated in the evaluation-performance relationship. They have found support for this account, using the antisaccade task (Jamieson & Harkins, 2007), GRE quantitative
problems (Jamieson & Harkins, 2009), and the Stroop (Jamieson & Harkins, 2011). The present research examines the effects of stereotype threat on sensorimotor performance and tests the predictions of the mere effort account.

6.2.4 Present Research and Hypotheses

Based on previous research on the ball bouncing task, predictions can be made for the effect of stereotype threat on the outcome variable and potential control mechanisms, both for novices and experienced participants. As noted above, in the virtual ball bouncing task, the initial tendency of novices is to hit the ball with positive acceleration. Adopting the perspective of the mere effort account, this behavior represents the prepotent or most likely response for novice performers. Hence, the mere effort account argues that stereotype threat should potentiate positive racket acceleration at impact in novices, leading to worse performance than controls. In contrast, experienced performers have learned to hit the ball with negative acceleration to exploit dynamic stability. Hence, hitting with negative acceleration is the prepotent response in experienced participants, who should therefore perform better than control participants under stereotype threat.

For novices, Schmader et al. (2008)’s working memory account would not appear to provide an a priori basis for predicting the effect of threat on acceleration or for acceleration’s effect on error. For experienced participants, Beilock’s explicit monitoring account predicts that stereotype threat would lead to debilitation, not facilitation, as threat leads them to think about enacting behaviors that have been proceduralized and run effectively without the contribution of working memory.

In the present research, we tested the predictions of the mere effort account for male and female novice and experienced performers using the rhythmic ball bouncing task. All participants were told that performance on this visuo-spatial task was highly
related to math ability, and that gender differences either had (stereotype threat) or had not (no stereotype threat) been found in performance on this task. This stereotype threat was intended to stigmatize the female participants, and thus potentiate their prepotent response.

In Experiments 1 and 2, novice participants performed 25 trials of the ball bouncing task; half of the male and female participants received the stereotype threat manipulation before starting the task. For these experiments on novice performers, we hypothesized that the prepotent response, hitting the ball with positive racket acceleration, would be potentiated in the stigmatized females. As a result, novice females under stereotype threat would perform with larger errors than control novices. We expected no performance differences between novice males under the same stereotype threat conditions.

In Experiment 3, female participants performed 12 trials of the ball bouncing task. At this point, a stereotype threat manipulation was implemented after which the participants performed an additional 12 trials. We hypothesized that in the initial 12 trials, the participants would learn to exploit dynamic stability such that in the second set of 12 trials the prepotent response would be to hit with negative racket acceleration. This response would then be potentiated in the females under stereotype threat, resulting in lower errors than control females.

### 6.3 Experiment 1

In Experiment 1, male and female participants were asked to perform 25 trials of the ball bouncing task, each lasting 40 sec. Prior to performing, participants were randomly assigned to either the stereotype threat or to the no-stereotype threat condition.
We hypothesized that stereotype threat would potentiate the prepotent response, hitting the ball with positive racket acceleration, making it more difficult for threatened females to hit the ball with negative acceleration. As performance would not utilize dynamic stability, errors from prior bounces would not be stabilized, and the threatened females should perform more poorly than non-threatened females. We further hypothesized that this effect would be absent in males.

6.3.1 Method

6.3.1.1 Participants

Seventy-two undergraduate students (36 males and 36 females) from Northeastern University participated in the experiment in exchange for partial fulfillment of a course requirement. None had any prior experience with the specific task.

6.3.1.2 Task

In the experimental task, the participant stood 2 m in front of rear projection screen (2.43 m x 2.43 m) holding a real table tennis racket in his or her dominant hand (Figure 6.1). A light rigid rod with two hinge joints was attached to the racket surface and ran through a wheel whose rotation was registered by an optical encoder at a sampling rate of ~500 Hz. While the joints allowed the racket to move and tilt with minimal friction in all three dimensions, the encoder only measured the vertical displacement of the racket. The position of the virtual racket on the screen, represented by a horizontal red line (0.2 m x 0.02 m), was controlled by the measured position of the real racket. The vertical position of the virtual ball, represented by a white filled circle (0.02 m radius), was determined using ballistic flight and inelastic instantaneous impact equations. Based on these equations, the maximum ball height of each bounce was determined by the ball velocity, racket velocity, and racket height at ball-racket impact. To simulate the haptic sensation
of a real ball hitting the racket, a mechanical brake was attached to the rod that was activated at ball-racket impact of each bounce. The participant was instructed to rhythmically bounce the ball to a target line, represented by a yellow horizontal line (1.0 m x 0.02 m) extending from the left edge to the middle of the screen, positioned 1 m above the minimum racket position.

6.3.1.3 Procedure

Each trial began with the ball appearing on the target line at the left side of the screen and rolling horizontally to the center of the screen. Upon reaching the center, the ball dropped vertically from the target line to the virtual racket. The participant was instructed by Experimenter 1 (male) to continuously bounce the ball to the exact vertical height of the target line for the duration of the trial. Each trial lasted 40 seconds and consisted of approximately 60-80 bounces.

All participants were given one practice trial for familiarization under the supervision of Experimenter 1 (male). After this practice trial, the experimenter informed the participants of the following: “The task you are about to complete is a test of visuospatial capacity. Performance on this task is closely linked to math ability. As you may know, there has been some controversy about whether there are gender differences in math and spatial ability. Previous research has demonstrated that gender differences exist on some of these tasks, but not on others. In our lab, we examine performance on both kinds of tasks.” Participants were then randomly assigned to a stereotype threat (ST) or no stereotype threat (N-ST) condition. ST participants were informed that the task had been shown to produce gender differences, whereas N-ST participants were told that it had not. This verbal instruction has been shown to produce stereotype threat effects in previous research (Brown & Pinel, 2003; Keller & Dauenheimer, 2003; O’Brien &
Crandall, 2003; S. J. Spencer et al., 1999). At this point, Experimenter 1 excused himself after introducing Experimenter 2 (female), who was blind to each participant’s condition. Experimenter 2 supervised each participant as she or he performed 25 trials (with a brief rest after 12 trials).

Following completion of 25 trials, the participants filled out a brief questionnaire for evaluating the effectiveness of the stereotype threat manipulation. The questionnaire asked: to what extent are there gender differences in performance on this task (1 = no gender differences and 11 = gender differences), who do you believe performs better on this task? (1 = males perform better, 6 = males and females perform the same, and 11 = females perform better), and to what extent is performance on this task related to mathematical ability (1 = not at all and 11 = closely related).

6.3.1.4 Data Reduction and Dependent Measures

Each bounce in the trial was defined as the event between two consecutive ball-racket impacts, or two consecutive ball position minima (Figure 6.4). Task performance was characterized by the median error of bounces in each trial. Error was defined as the unsigned difference between the target height and the maximum ball amplitude for each bounce in the trial. Median racket acceleration at impact of each trial was used as a criterion whether participants performed with dynamically stable solutions. To obtain acceleration of the racket trajectory, the racket position was resampled at a constant frequency of 1 kHz and double-differentiated using a second-order Savitzky-Golay filter with a window size of .05 s (Savitzky & Golay, 1964). Racket acceleration at impact was defined as the racket acceleration 20 ms prior to the first minimum of ball position at each bounce. The interval of 20 ms was chosen to avoid capturing any artifacts due to the activation of the mechanical brake.
To identify the presence of active error corrections by the participants, it was necessary to tease apart the automatic error stabilization due to dynamic stability from such additional active error-based corrections. To do so, we followed the approach detailed in (K. Wei et al., 2008) that quantified the self-stabilizing properties of the map with the autocorrelation function. The autocorrelation of successive ball release velocities at impact determined how fast errors dissipate, or exponentially decay, across impacts. The correlation values of the function relate to the time constant of exponential error decay over bounces. Positive autocorrelation values correspond to exponential decay, with faster decay indicated by smaller positive values. However a simple autocorrelation analysis of bounce-to-bounce fluctuations was not sufficient as errors also decline as a function of dynamic stability. To distinguish between error decreases due to dynamic stability and active corrections, K. Wei et al. (2008) assessed the error dynamics in a ball bouncing model with an added stochastic component. As this model did not include any error correction, the autocorrelation values quantified the amount of error dissipation due to dynamic stability. K. Wei et al. (2008) concluded that if the lag-1 autocorrelation values obtained from participant performance were more negative than those generated by the stochastic model, then the participant applied active error corrections to decrease the error faster than with error stabilization alone.

Comparison of the lag-1 autocorrelation value from the stochastic model with the autocorrelation in participant performance could therefore quantify the relative amount of active error correction. Using the same parameters from the virtual task in the stochastic model rendered a lag-1 autocorrelation value of 0.24. Therefore a lag-1 autocorrelation
value from participant performance below 0.24 indicated the presence of active error correction in a given trial.

6.3.1.5 Statistical Analysis of Performance

All performance measures were analyzed in 2 (Threat) x 2 (Gender) x 25 (Trials) analyses of variance (ANOVAs) with threat and gender as between-subjects factors and trials as a within-subjects factor. The Greenhouse-Geisser correction factor was applied to the within-subject effects (Kirk, 1995). Prior to these analyses, all dependent measures were checked for deviations from normal distributions using W-tests (Shapiro & Wilk, 1965). Medians of absolute error and racket acceleration were used because their deviations within each trial were not strictly Gaussian.

6.3.2 Results

6.3.2.1 Perception of Stereotype Threat

The measures of perception of threat obtained from the questionnaires were analyzed in 2 (Threat) x 2 (Gender) between-subjects ANOVAs. Participants in the ST condition reported that gender differences existed in this task to a greater extent ($M = 7.11, SD = 2.63$) than in the N-ST condition ($M = 3.43, SD = 2.81$), $F(1, 68) = 34.86, p < .001, d = 1.43$. Neither the main effect for gender, nor the interaction was significant, $Fs < 1, ps > .30$.

ST participants also tended to report that males performed better on the task ($M = 4.35, SD = 1.96$) compared to N-ST participants ($M = 4.97, SD = 1.38$), $F(1, 68) = 2.65, p = .11, d = .39$. Neither the main effect for gender, nor the interaction was significant, $Fs < 1, ps > .40$. Although the threat main effect was only marginal, the same manipulation check was used in each of the three experiments reported in this work. Taken together across the three experiments, the effect was highly reliable with a combined $z = 3.98, p <$
.0001, $d = .64$. The effect for the gender difference question was also highly reliable across the three experiments, combined $z = 6.78, p < .00001, d = 1.41$.

Analysis of the question asking how related performance on the task was to mathematical ability revealed no significant differences, $F_s < 1, ps > .70$. The grand mean on this measure ($M = 6.96, SD = 2.62$) was significantly different from the midpoint of the scale (6), $t(71) = 3.10, p < .01, d = .37$.

### 6.3.2.2 Performance Error

The primary performance measure on the racket task was the median absolute error in each trial for each participant. Given that the target height determined the number of bounces for a given trial duration, each trial had on average 50 bounces. The 2 (Threat) x 2 (Gender) x 25 (Trial) ANOVA on the median error revealed a significant main effect for trial, $F(24, 1632) = 15.21, p < .0001$, reflecting the fact that the magnitude of error dropped over the course of the 25 trials from 26 cm ($SD = 23$ cm) in trial 1 to 8 cm ($SD = 10$ cm) in trial 25 (Figure 6.5A). The analysis also revealed a marginal main effect for threat, $F(1, 68) = 2.74, p < .11, d = .40$. Participants in the ST condition tended to exhibit larger errors across the 25 trials ($M = 13$ cm, $SD = 14$ cm) than participants in the N-ST condition ($M = 10$ cm, $SD = 11$ cm). Neither the main effect for gender nor the Gender x Threat interaction reached significance, $F_s < 1, ps > .70$. The Trial x Gender, $F(24, 1632) = 1.61, p = .15$ and the three-way interaction, $F < 1, p > .50$, were also not significant.

### 6.3.2.3 Racket Acceleration at Impact

Racket acceleration at impact served as the measure of dynamic stability in task performance. As observed in previous studies, participants initially hit the ball with positive racket accelerations until they learned to exploit the dynamically stable solution that produced automatic stabilization for small errors. Consistent with these results, the
median racket acceleration at impact again revealed a significant trial effect, \( F(24, 1632) = 38.10, p < .0001 \), indicating that the racket acceleration values changed from positive in trial 1 (\( M = 3.34 \text{ m/s}^2, SD = 3.81 \text{ m/s}^2 \)) to negative values in trial 25 (\( M = -1.16 \text{ m/s}^2, SD = 2.90 \text{ m/s}^2 \)) (Figure 6.5B). Analysis of this measure also revealed a main effect for gender, \( F(1, 68) = 4.47, p < .05, d = .51 \). Overall, females tended to hit the ball with more negative racket acceleration (\( M = -0.85 \text{ m/s}^2, SD = 3.28 \text{ m/s}^2 \)) than males (\( M = 0.32 \text{ m/s}^2, SD = 3.11 \text{ m/s}^2 \)). Neither the threat main effect nor the Gender x Threat interaction was significant, \( Fs < 1, ps > .40 \). The interaction for Gender x Trial was significant, \( F(24, 1632) = 2.43, p < .02 \), as females hit the ball with greater positive racket acceleration in trial 1 (\( M = 3.78 \text{ m/s}^2, SD = 3.58 \text{ m/s}^2 \)) than males (\( M = 2.89 \text{ m/s}^2, SD = 4.02 \text{ m/s}^2 \), \( F(1, 1632) = 4.26, p < .05 \), but ended the experiment with lower acceleration values, showing more improvement (e.g., trial 25: \( F(1, 1632) = 10.91, p < .05 \); females: \( M = -1.88 \text{ m/s}^2, SD = 2.57 \text{ m/s}^2 \); males: \( M = -0.44 \text{ m/s}^2, SD = 3.07 \text{ m/s}^2 \)). Neither the Trial x Threat, \( F < 1, p > .80 \), nor the three-way interaction was significant, \( F(24, 1632) = 1.08, p > .35 \).

6.3.2.4 Lag-1 Autocorrelation of Release Ball Velocities at Impact

Lag-1 autocorrelation of release ball velocities at successive impacts measured the presence of error stabilization and active error correction. A \( t \)-test revealed that the mean lag-1 autocorrelation value of all participants was positive (\( M = .13, SD = .11 \)), consistent with results from previous analysis (K. Wei et al., 2008). Importantly, the autocorrelation values for both males and females were significantly lower than 0.24, the value obtained from stabilization in the stochastic ball bouncing model without error correction. This suggested that additional control by participants reduced the correlations and suggested active error corrections. ANOVA revealed that lag-1 autocorrelation significantly decreased across trials, \( F(24, 1632) = 6.83, p < .0001 \), from .32 (\( SD = .29 \)) in trial 1 to .07
Further, the main effect for gender was significant, $F(1, 68) = 8.53, p < .05, d = .71$. While the mean autocorrelation values for males and females were both significantly less than 0.24, males used significantly more error correction ($M = .09, SD = .23$) than females ($M = .17, SD = .28$). There was no main effect for threat, $F < 1, p > .65$, nor was there a Gender x Threat interaction, $F < 1, p > .45$. The Trial x Threat, $F < 1, p > .90$, the Trial x Gender, $F(24, 1632) = 1.31, p > .18$, and the three-way interaction, $F(24, 1632) = 1.05, p > .39$, were also not significant.

### 6.3.3 Discussion

All participants learned to perform the task across practice as expected from previous results (de Rugy et al., 2003; Schaal et al., 1996; K. Wei et al., 2008). Both males and females showed similar improvements in the primary performance measure, median error. A marginal main effect for threat was seen as participants in the stereotype threat conditions tended to perform more poorly than their unthreatened counterparts. However, this effect was present in both males and females, counter to the hypotheses. Moreover, there was no hint of an effect of stereotype threat on racket acceleration at impact, the measure of dynamic stability.

One possible reason why we did not see the hypothesized Gender by Threat interaction is that we underestimated the effect of active error correction on task performance. Based on past research we expected active correction to play little, if any, role in performance (K. Wei et al., 2007). Instead, over trials, we found decreases in both racket acceleration values and the measure of active error correction, the lag-1 autocorrelations of successive ball release velocities, suggesting that participants learned to employ both active correction and dynamic stability to improve their performance. The
findings also suggest that males relied on active correction more than dynamic stability, whereas females relied more on dynamic stability than on active correction.

It is possible that the complex interplay of the two strategies accounts for the absence of the predicted stereotype threat effect. Ehrlenspiel et al. (2010) faced a similar problem in previous research using the same task. In their first experiment, participants learned to perform the ball bouncing task over the course of 32 trials on the first day. On the second day, the participants were randomly assigned to a high-stress or to a no-stress group. The high-stress participants were told that they had been entered in a competition with another participant. The participants then performed another 32 trials of the task. Ehrlenspiel et al. (2010) found that the participants subject to performance pressure improved more from day 1 to day 2 than participants in the control group. This finding is consistent with the mere effort account in that the participants learned to perform the task on day 1. As a result, the correct responses probably became more likely and, potentiated by the performance pressure, produced better performance in the competition condition. However, just as in our experiment, there were no differences in the measures of active control and dynamic stability. These differences were likely absent because participants may have adopted different approaches to error reduction, similar to our experiment.

In the ball bouncing task, it is not possible to identify one trial as generated by active correction and another by dynamic stability, as both strategies are simultaneously employed. Thus, in a second experiment, Ehrlenspiel et al. (2010) introduced perturbations to the flight of the ball on all trials of Day 2. These perturbations were too large to be self-correcting due to dynamic stability. This effectively minimized the contribution of dynamic stability to error reduction and made active error correction the
only possible solution for reducing error. The results still showed improved performance in the high-stress condition, but there was also evidence that this group used active control more than the no-stress group. We conclude that this increase in error correction in the modified task was possible, because other potential compensatory mechanisms provided by dynamic stability were absent.

Thus, Experiment 2 introduced a manipulation that prevented participants from actively correcting the observed error in the previous bounce. We hypothesized that removing this compensatory mechanism would better reveal the effect of stereotype threat on task performance.

6.4 Experiment 2

Similar to Experiment 1, male and female participants were asked to perform 25 trials of the ball bouncing task. Again, participants were randomly assigned to either the stereotype threat or the no-stereotype threat condition prior to performing the task. In this experiment, however, unbeknownst to the participants, we applied a time-dependent manipulation to the racket velocity at ball-racket impacts as described in Huber and Sternad (2015). Instead of using the actual racket velocity at ball-racket impact to determine ball position for the virtual display, we used the racket velocity 25 samples (50 ms) prior to impact (see Appendix B). This time shift of the racket velocity altered the mapping between the perceived error and the corrective action, making actions that were normally successful in correcting an error no longer effective. If the racket impacted the ball in the accelerating portion of the trajectory, the manipulation caused errors to propagate faster and performance became unstable. Because hitting the ball with positive acceleration led to higher errors that could not be actively corrected, the only way to
successfully perform the task was to exploit dynamic stability by hitting the ball with negative racket acceleration.

Huber and Sternad (2015) showed that there was no significant difference in error between participants who performed the task with and without the time-dependent manipulation, indicating that the manipulation did not change the difficulty of the task. Furthermore, they found that even in this manipulated task, the median racket acceleration at impact was initially positive. This indicated that the prepotent response in the task with the time shift was the same as under previous conditions. Thus, we maintained the hypotheses from Experiment 1: stereotype threat potentiates the response of hitting with positive racket acceleration. As a result, the threatened females should perform more poorly and larger errors were expected. In contrast, threatened males should be unaffected by the threat. We expected to see this result more clearly than in Experiment 1, because it was necessary to hit the ball with negative racket acceleration to achieve error stabilization.

6.4.1 Methods

6.4.1.1 Participants

Sixty-nine Northeastern University undergraduate students (32 males and 37 females) participated in the experiment in exchange for partial fulfillment of a course requirement. None had any prior experience with the specific task.

6.4.1.2 Task and Procedure

The task and procedure was identical to that described in Experiment 1, including all manipulations and questionnaire items. However, the calculation and simulation of the ball trajectory did not use the veridical racket velocity at ball-racket impact, but the racket velocity 25 samples (50 ms) prior to impact.
6.4.1.3 Data Reduction and Dependent Measures

All data analysis and reduction measures were identical to Experiment 1 with two exceptions: 1) In the perturbed case, active error correction could no longer be determined and was excluded from the data analysis in this experiment; and 2) we needed to eliminate short portions of the trials. Due to the manipulation, the ball occasionally became unstable and “stuck” on the racket as the participant moved the racket up and down. During these “stuck bounces,” the ball did not reach the target line and the errors and other dependent measures were very large. Including these values in the average estimates of performance would have significantly skewed these measures, without adding to our understanding of how participants performed the task. Consequently, these uncharacteristic bounces were excluded from each trial before calculating the dependent measures. To identify these events, an investigator who was blind to the experimental condition defined a threshold distance of 0.25m between the maximum ball position and maximum racket position; bounces that were below this threshold were eliminated. Participants with a very high number of stuck bounces across trials were omitted from further analysis, as it was impossible to accurately assess their performance. The criterion for elimination of participants was when the mean duration of stuck bounces across trials was longer than two standard deviations away from the overall participant mean.

6.4.1.4 Statistical Analyses

The performance measures were analyzed in 2 (Threat) x 2 (Gender) x 25 (Trials) ANOVAs, with threat and gender as between-subjects factors and trials as a within-subjects factor.
6.4.2 Results

6.4.2.1 Perception of Stereotype Threat

The perception of threat assessed in questionnaires was analyzed in 2 (Threat) x 2 (Gender) between-subjects ANOVAs. Participants in the ST condition reported that gender differences existed in this task to a greater extent ($M = 6.79, SD = 2.78$) than N-ST participants ($M = 3.97, SD = 2.93$), $F(1, 65) = 16.29, p < .001, d = 1.00$. Participants in the ST condition also reported that males performed better on the task to a greater extent ($M = 4.27, SD = 1.75$) than N-ST participants ($M = 5.29, SD = 1.25$), $F(1, 65) = 7.48, p < .01, d = .69$. Neither the gender main effect nor the Gender x Threat interaction was significant, $F$s < 1, $p$s > .80.

Once again, analysis of the question asking how related performance on the task was to mathematical ability revealed no significant differences, $F$s < 1, $p$s > .40. The grand mean on this measure ($M = 7.46, SD = 2.10$) was significantly different from the midpoint of the scale ($6$), $t(68) = 5.78, p < .0001, d = .70$. This pattern of findings shows that the stereotype threat manipulation was successfully implemented.

6.4.2.2 Failed Performance

Applying the exclusion criterion based on mean duration of stuck bounces, four participants (two males and two females) were excluded from further analysis. They exhibited an average of 13.5 to 20 s per 40-s trial, where the ball was close to the racket and never achieved regular rhythmic behavior. This substantially exceeded the overall mean of 2.94 s per trial. An analysis performed on the mean duration of stuck bounces across trials for the remaining 65 participants revealed no reliable group differences, $p$s > .16. The analysis yielded a significant main effect for trial, $F(24, 1464) = 14.40, p <$
Across trials, the amount of time spent in this behavior dropped from an average of 6.87 s in trial 1 to 0.73 s in trial 25.

### 6.4.2.3 Performance Error

Analysis of the performance error revealed a significant change across trials, $F(24, 1464) = 29.42, p < .0001$, reflecting the fact that the participants’ performances improved from trial 1 ($M = 21$ cm, $SD = 9$ cm) to trial 25 ($M = 10$ cm, $SD = 7$ cm). Participants reached an asymptote at about trial 18 (Figure 6.6A). This analysis also revealed a reliable Gender x Threat interaction, $F(1, 61) = 4.24, p < .05, d = .53$. A planned contrast showed that the error for ST females was greater ($M = 17$ cm, $SD = 8$ cm) than for N-ST females ($M = 12$ cm, $SD = 6$ cm), $F(1, 61) = 7.35, p < .01, d = .69$. In contrast, the performance of ST males ($M = 11$ cm, $SD = 6$ cm) and N-ST males ($M = 12$ cm, $SD = 7$ cm) did not differ, $p > .50$ (Figure 6.6B). The main effect for gender, $F(1, 61) = 6.32, p < .05, d = .64$, and the marginal threat main effect, $F(1, 61) = 2.65, p = .11, d = .42$, must be interpreted in the context of the two-way interaction, suggesting that all participants learned the task, but the threatened females started out and remained at a lower skill level than the others. The Gender x Trial, Threat x Trial, and three-way interaction were all nonsignificant, $Fs < 1, ps > .45$.

### 6.4.2.4 Racket Acceleration at Impact

Analysis of the median racket acceleration at impact revealed a significant trial effect, $F(24, 1464) = 34.67, p < .0001$, which reflected the fact that the participants’ racket accelerations at impact went from positive in trial 1 ($M = 5.02$ m/s$^2$, $SD = 4.75$ m/s$^2$) to negative in trial 25 ($M = -3.24$ m, $SD = 2.59$ m), reaching an asymptote at about trial 20 (Figure 6.6C). This analysis also produced a Gender x Threat interaction, $F(1, 61) = 8.39, p < .01, d = .74$. The average racket acceleration at impact across the 25 trials for
threatened females was positive ($M = .54 \text{ m}, SD = 5.23 \text{ m/s}^2$), whereas it was negative for N-ST females ($M = -1.67 \text{ m/s}^2, SD = 3.88 \text{ m/s}^2$), $F(1, 61) = 7.28, p < .01, d = .69$ (Figure 6.6D). While both threatened and unthreatened females learned to hit with negative acceleration by the end of the experiment, threatened females hit with less negative racket acceleration throughout, as shown in Figure 6.6D. The fact that they hit with positive acceleration for approximately the first ten trials resulted in a positive mean overall. Racket acceleration at impact for ST males ($M = -1.68 \text{ m/s}^2, SD = 3.51 \text{ m/s}^2$) did not differ from N-ST males ($M = -0.41 \text{ m/s}^2, SD = 4.28 \text{ m/s}^2$), $p > .15$. The main effects for threat and gender, the two-way interactions, and the three-way interaction were all non-significant, $Fs < 1, ps > .40$.

6.4.3 Discussion

As in Experiment 1, the participants’ performance improved over the course of the 25 trials, which was paralleled by a decrease in racket acceleration at impact towards more negative values. Like Experiment 1, the performance of threatened females was significantly worse than that of non-threatened females. Unlike in Experiment 1, however, threatened females were less successful in exploiting dynamic stability than non-threatened females, consistent with the main hypothesis. In fact, the average racket acceleration at impact across all 25 trials was positive for threatened females, whereas it was negative for non-threatened females. For males, there was no difference in error or acceleration as a function of threat. The marginal debilitation effect found in the male/threat condition in Experiment 1 did not recur.

By applying the time-shift manipulation, we were able to see that threatened females had more difficulty performing the task than non-threatened females. Moreover, we were able to see that this difficulty stemmed from the fact that, at the outset, hitting
the ball with positive racket acceleration was the prepotent response, and that threat potentiated this response. These findings are consistent with the mere effort account in that threat debilitates performance when the prepotent response is incorrect.

The mere effort account also predicts that when threat potentiates a prepotent response that is correct, performance is improved. This means that if the participants had already learned to exploit dynamic stability, the response of hitting the ball with negative racket acceleration would be prepotent, and should be potentiated by threat. As a result, threatened females would be expected to perform better, not worse, than non-threatened females. This hypothesis was tested in Experiment 3.

6.5 Experiment 3

In this experiment, we recruited only female participants, who performed 24 trials of the ball bouncing task with the same manipulation introduced in Experiment 2. However, the first block of 12 trials was performed without any threat manipulation. Examination of the error results in Experiment 2 had suggested that by trial 12 participants had learned to hit the ball with negative acceleration, and thereby acquired a certain level of expertise. At this point, the females were randomly assigned to a threat or a no-threat condition, using the same manipulation as that employed in Experiment 2. To the extent that the prepotent response was now to hit the ball with negative acceleration and exploit dynamic stability, we hypothesized that threat would potentiate this correct response, leading to a better performance by threatened females than their non-threatened counterparts. Thus, unlike Experiment 2, females subject to stereotype threat were hypothesized to perform better, not worse, than non-threatened females.
6.5.1 Methods

6.5.1.1 Participants

Thirty-five Northeastern University undergraduate females participated in the experiment in exchange for partial fulfillment of a course requirement. None had any prior experience with the specific task.

6.5.1.2 Task and Procedure

The task and procedure were identical to that described in Experiment 2, including all manipulations and questionnaire items. The only difference was that the threat manipulation was implemented after 12 trials instead of at trial 1 and the experiment continued for another 12 trials. At the outset, a male experimenter described the ball bouncing procedure followed by a practice trial. At this point, he left the room and a female experimenter collected the data for the first 12 trials. She then indicated that she needed to take a brief break, and the male experimenter returned. He then implemented the stereotype threat manipulation under the guise of providing some background information on the research, while they waited for the other experimenter to return. The female experimenter then returned and was blind to the experimental condition as she collected the data for the last 12 trials.

6.5.1.3 Data Reduction and Dependent Measures

The method of data reduction, participant exclusion criteria, and dependent measures were identical to those described in Experiment 2.

6.5.1.4 Statistical Analysis

Experiment 3 was conducted in two 12-trial blocks, separated by the stereotype threat manipulation that was given after the first block. For each of the performance
measures, we conducted a 2 (Threat) x 12 (Trials) ANOVA, one for Block 1 and one for Block 2.

6.5.2 Results

6.5.2.1 Perception of Stereotype Threat

The perception of threat measures were analyzed in a one-way ANOVA with Threat as the independent variable. Females in the ST condition reported that gender differences existed to a greater extent ($M = 6.06, SD = 3.19$) than N-ST females ($M = 2.82, SD = 1.93$), $F(1, 33) = 12.48, p < .01, d = 1.23$. Females in the ST condition also reported that males performed better on the task to a greater extent ($M = 4.11, SD = 1.94$) than N-ST participants ($M = 5.65, SD = 1.27$), $F(1, 33) = 7.59, p < .01, d = .96$.

Analysis of the question asking how related performance on the task was to mathematical ability revealed no significant differences, $p > .50$. The grand mean on this measure ($M = 7.43, SD = 2.51$) was significantly different from the midpoint of the scale (6), $t(34) = 3.36, p < .01, d = .57$. Once again, this pattern of findings showed that stereotype threat was successfully implemented.

6.5.2.2 Failed Performance

First, the intervals with stuck bounces were determined in both blocks. On the basis of performance in Block 1, one participant was eliminated (average of 19.08 s of stuck bounces per trial vs. overall $M = 6.14$ s per trial). For the remaining 34 participants, the mean duration of stuck bounces across trials in Block 1 revealed a reliable trial effect, $F(11, 352) = 18.77, p < .0001$. The duration of stuck bounces dropped from an average of 15.71 s in trial 1 to 2.79 s in trial 12. There was no reliable threat effect, nor an interaction, $ps > .30$. In Block 2, the analysis of stuck bounces revealed no group
difference, \( p = .19 \), no trial effect, \( p = .19 \), nor their interaction, \( p > .40 \). The participants spent 2.18 s in stuck bounces in the first trial of the second block and .49 s in the last.

### 6.5.2.3 Performance Error

Analysis of the first block of 12 trials produced a reliable main effect for trial, \( F(11, 352) = 20.36, p < .0001 \). As shown in Figure 6.7A, the average median error dropped from 29 cm (\( SD = 10 \) cm) in trial 1 to 14 cm (\( SD = 8 \) cm) in trial 12. There was neither a threat effect, \( F(1, 32) = 1.09, p > .30 \) (Figure 6.7B), nor an interaction, \( F(1, 32) = 1.03, p > .40 \).

The second block of trials also showed a reliable trial effect, \( F(11, 352) = 4.63, p < .001 \). The errors dropped from a mean of 14 cm (\( SD = 7 \) cm) in trial 1 to 10 cm (\( SD = 5 \) cm) in trial 12 of Block 2. This analysis also revealed a reliable threat effect, \( F(1, 32) = 5.48, p < .05, d = .83 \). Females under stereotype threat performed better (\( M = 9 \) cm, \( SD = 4 \) cm) than their non-threatened counterparts (\( M = 13 \) cm, \( SD = 7 \) cm) (Figure 6.7B). The interaction of these variables was not significant, \( F < 1, p > .60 \).

### 6.5.2.4 Racket Acceleration at Impact

In Block 1, there was a reliable trial effect, \( F(11, 352) = 10.32, p < .0001 \). As shown in Figure 6.7C, median racket acceleration at impact dropped from a mean of 4.19 m/s\(^2\) (\( SD = 5.86 \) m/s\(^2\)) in trial 1 to a mean of -1.26 m/s\(^2\) (\( SD = 4.11 \) m/s\(^2\)) in trial 12. Neither the threat main effect, \( F(1, 32) = 1.03, p > .30 \) (Figure 6.7D), nor the interaction was significant, \( F < 1, p > .65 \).

In Block 2, once again the trial main effect was reliable, \( F(11, 352) = 5.69, p < .0001 \), with the mean falling from -0.38 m/s\(^2\) (\( SD = 5.07 \) m/s\(^2\)) in trial 1 to -3.08 m/s\(^2\) (\( SD = 2.78 \) m/s\(^2\)) in trial 12. This analysis also revealed a significant stereotype threat main effect, \( F(1, 32) = 4.70, p < .05, d = .77 \). Threatened females hit the virtual ball with
greater negative racket acceleration \((M = -3.50 \text{ m/s}^2, SD = 3.11 \text{ m/s}^2)\) than non-threatened females \((M = -1.81 \text{ m/s}^2, SD = 3.07 \text{ m/s}^2)\) (Figure 6.7D). The interaction was not significant, \(F < 1, p > .50\).

### 6.5.3 Discussion

The mere effort account suggests that social threat potentiates prepotent responses. If those responses are incorrect, threat debilitates performance as we found in Experiment 2. At the outset of the task the prepotent response was hitting the ball with positive acceleration, which was potentiated by threat, leading to poorer performance by threatened females than by non-threatened ones. However, in Experiment 3, females learned to hit the ball with negative acceleration prior to the implementation of the threat manipulation. As a result, the prepotent response was to hit with negative acceleration, and the potentiation of this response to exploit dynamic stability led to better performance by threatened participants than by non-threatened ones as hypothesized.

### 6.6 General Discussion

The effect of stereotype threat on the performance of stigmatized individuals has been studied in a variety of cognitive tasks. However, only a few studies have examined the effect of stereotype threat on sensorimotor performance, and in these studies performance has been quantified with only relatively coarse measures (e.g., distance from the ball to the target after a putt; the time needed to dribble through a slalom course). While such performance measures characterize overall performance, they do not capture more fine-grained aspects of the execution that may shed light on the mechanisms of control through which stereotype threat has its effects on performance.
Harkins (2006) previously raised a similar concern with regard to cognitive tasks, arguing that a comprehensive understanding of the task is necessary to gain insight into how social threat affects task performance. In fact, it was the in-depth analysis of the Remote Associates Task that led to the formulation of the mere effort account for stereotype threat effects (Jamieson & Harkins, 2007). The mere effort account suggests that participants do not simply fall victim to a process that negatively affects performance. Instead, it suggests that stereotyped participants actually intensify their efforts during task performance, but these efforts may be misdirected. The mere effort account argues that under stereotype threat the response that is most likely for the task is potentiated. If this response is correct, then performance will be improved. If incorrect, and the participant does not recognize this, or does not have the time or skill necessary for correction, performance will be debilitated. However, identifying the prepotent response at the control level, and its role in overall task performance requires a thorough understanding of the task demands.

In this study, we chose the sensorimotor skill of rhythmically bouncing a ball to a target height, because previous theoretical analysis of the task dynamics provided the foundation for teasing out the effect of threat on motor control (Schaal et al., 1996; Sternad, Duarte, et al., 2000). Whereas prior psychological studies used outcome measures, such as error, to draw inferences about the effect of stereotype threat, the ball bouncing task afforded more direct assessment of control-relevant variables. In the ball bouncing task, novice participants initially hit the ball with positive acceleration, and then learn to stabilize error by shifting ball-racket impact from positive to negative acceleration. Thus, the mere effort account predicts that performance should be
debilitated for stigmatized novices, as their most likely (prepotent) response is to hit with positive acceleration. We tested this hypothesis in Experiment 1. Counter to expectation, the results showed no significant difference between threatened and non-threatened females in the control-relevant variable, racket acceleration at impact. Further, threat had a tendency to debilitate the performance of both males and females in the stereotype threat condition. We attributed this finding to the fact that active error correction may have confounded results and complicated inferences about control and stereotype threat.

While exploitation of dynamic stability is a signature of expert performance, other control processes, such as active error correction can also be used to reduce error (de Rugy et al., 2003; Ronsse & Sternad, 2010; Siegler et al., 2010; K. Wei et al., 2007, 2008). In Experiment 1 we found that males and females used varying degrees of dynamic stability and active correction to achieve the same level of task performance. The complex interplay of these strategies to reduce error may account for the absence of the predicted effects. Had we only analyzed error, we would not have known that the interplay of the control mechanisms likely masked the effect of threat on error.

The results of Chalabaev, Sarrazin, et al. (2013) also illustrate the need to consider the effect of ST on other measures beyond overall performance. As their isometric force task was a simple one-step skill, the null effect of ST in the primary performance measure alone could lead to the interpretation that non-proceduralized tasks, which are not susceptible to explicit monitoring processes, are immune to the effects of stereotype threat. Instead, by considering a secondary measure, Chalabaev, Sarrazin, et al. (2013) revealed that ST did affect motor control even in the absence of explicit monitoring processes. These findings provide compelling evidence that outcome performance
measures alone are not sufficient to shed light on the mechanisms through which stereotype threat affects motor performance.

Whereas Chalabaev, Sarrazin, et al. (2013) were able to capture the effect of ST by a second measure of performance, the complex ball bouncing task did not permit such a simple parsing of variables. Thus, we applied the time-shift manipulation with the goal of isolating the effect of ST on one of the control mechanisms. In earlier research on ball bouncing under stress, Ehrlenspiel et al. (2010) faced a similar problem in that they observed an effect of stress on error, but no effects on the mechanisms of control. In that research, as in Experiment 1, it is possible that some participants used active control, whereas others tended to rely on dynamic stability as a control mechanism. In a second experiment, Ehrlenspiel et al. (2010) therefore added random perturbations to the ball flight, which required the participants to use active error corrections and ruled out sole reliance on dynamic stability. This experimental variation led to the expected relationship between active correction and error.

Therefore, this study added a time-dependent manipulation to the racket velocity at impact in Experiments 2 and 3, which required participants to exploit dynamic stability to perform successfully. Using the modified task, Experiment 2 tested the initial hypothesis that performance by threatened novices would be degraded, because the prepotent response, hitting with positive acceleration, was potentiated. Consistent with this hypothesis, we found that the threatened novice females hit the ball with more positive racket accelerations and had higher errors than their unthreatened counterparts.

In Experiment 3, the stereotype threat manipulation was delivered after the females had practiced the task for 12 trials; having gained a moderate level of expertise, it was
expected that the prepotent response would now be to hit with negative acceleration. We hypothesized that this response would be potentiated in experienced females stigmatized by stereotype threat, and thus facilitate their performance. Indeed, threatened females hit the ball with greater negative acceleration than non-threatened females, which was accompanied by smaller errors. A caveat is that we showed this overt effect only in the modified task. However, we infer that similar processes hold for the unmanipulated task, only masked by compensatory processes due to error correction.

The effect of stereotype threat on task performance has been conceptualized in two lines of work. With a focus on cognitive tasks, the model of Schmader et al. (2008) argues that the effect of stereotype threat on cognition, affect, and motivation combine to impair the efficiency of working memory. This model also has a pathway that incorporates findings of Beilock and colleagues on sensorimotor tasks (Beilock et al., 2006). Thus, it proposes that the performance of experienced participants is disrupted by stereotype threat, because the threat leads them to monitor processes that have been proceduralized (i.e., well-learned). A second line of theorizing incorporates the same two processes, but has focused primarily on the effect of threat on the performance of sensorimotor tasks by experienced participants (DeCaro et al., 2011).

In accounting for the present results, one could argue that novices use working memory as they learn to perform the task, and that stereotype threat creates anxiety and self-doubt, which take up processing capacity, leading to the performance debilitation found in Experiment 2. For example, Stone et al. (1999) found that stereotype threat debilitated the performances of novice golfers. However, the working memory account provides no a priori basis for predicting the effect of threat on performance, nor does this
account provide any basis for expecting racket acceleration to mediate the effect. In contrast, based on identification of the prepotent or most probable response, the mere effort account makes an a priori prediction for the effect that stereotype threat will have on the control mechanism and the terminal behavior.

For experienced participants, Beilock’s explicit monitoring account predicts that threat would debilitate performance. For example, DeCaro et al. (2011) argue that in Beilock et al. (2006)’s golf putting task and in Chalabaev et al. (2008)’s soccer dribbling task expert performance was debilitated because stereotype threat led these participants to monitor their well-learned behaviors. However, in Ehrlenspiel et al. (2010) and in Experiment 3 of the current research, participants had learned to perform the task, which facilitated, not debilitated performance. Ehrlenspiel et al. (2010) suggested that this facilitated performance of their threatened participants may have been the result of the rhythmic nature of the task, which might make it less susceptible to the effects of explicit monitoring. The same caveat may also hold for the current experiments.

As mentioned in Ehrlenspiel et al. (2010), a previous brain imaging study (Schaal et al., 2004), behavioral results (Howard et al., 2011; Ikegami et al., 2010; Sternad et al., 2013), and modeling studies (Ronsse et al., 2009; Sternad, Dean, & Newell, 2000) support the argument that different control strategies are used to perform rhythmic and discrete motor tasks. Hence, the influence of stereotype threat on these different control strategies could potentially lead to different performance effects in the discrete golf putting task (Beilock et al., 2006). However, this does not account for the debilitation observed in the soccer dribbling task (Chalabaev et al., 2008). The soccer dribbling task is a continuous, and possibly rhythmic task, similar to ball bouncing. The essential
element in such continuous tasks is that errors continuously propagate from one ball contact to the next. In golf putting on the other hand, each putt is a new event with different initial conditions. While it is still unclear from these collective findings whether stereotype threat produces different effects in discrete and rhythmic tasks, it is an important consideration for future studies involving sensorimotor tasks.

Another potential reason for these divergent results could be that participants in these studies were not experienced in the specific experimental tasks. For example, in Beilock et al. (2006)’s research on golf putting, they pointed out that their “experts” were asked to putt the ball so that it stopped directly on the target. This specific experimental task is different from the task on which these participants are actually expert: putting the ball through the target into the hole. Beilock et al. (2006) acknowledged this fact, noting that experts found that the constraint of stopping the ball on the target made the task difficult. Similarly, the experimental task used in the soccer study by Stone and McWhinnie (2008) required participants to dribble with only the dominant foot, whereas they typically dribble with both feet. Thus, in neither case was the experimental task actually the task in which the participants were expert.

The detailed understanding of the ball bouncing task provides a theoretically grounded measure of expertise. In Experiments 1 and 2, the mean racket acceleration at impact in trial 1 was positive, assuring that the participants were indeed novices. In Experiment 3, participants were trained on the experimental task until they were experienced as indicated by negative racket acceleration at impact by trial 12 (Figure 6.7). Thus, the measure that determined the level of expertise was the same control mechanism that accounted for error performance on the specific ball bouncing task that
we used. These considerations demonstrate how a fine-grained task analysis can contribute to a better understanding of how a psychological manipulation (e.g., stereotype threat) affects a behavioral measure.

These findings also highlight the fact that verbal task instructions to the participant can subtly affect motor learning and performance. Previously, different types of instructions have been shown to affect complex skill learning, most notably in terms of directing attentional focus (Perkins-Ceccato, Passmore, & Lee, 2003). These more subtle effects during uncontrolled, spontaneous interaction with the subject are generally ignored in motor control research. On the other hand, gender differences have been documented, particularly that boys have better motor and visuo-spatial abilities than girls (Müller & Sternad, 2004a). More recently, modeling of brain activation networks has suggested a physiological basis for this gender difference (Ingalhalikar et al., 2014). In motor neuroscience these seemingly subtle effects of stereotype and gender on motor performance have been given little attention. Our results suggest that experimenters should take extra precautions to ensure that they do not indirectly elicit this stereotype in interactions with their participants to avoid confounding effects on experimental results.

Even if experimenters take such precautions in their interactions with their participants, it is still possible for subtle cues in the experimental setting itself to activate gender stereotypes. For example, research has shown that solo status (female tested by male experimenter along with other male participants) can produce threat effects (Minzlicht & Ben-Zeev, 2000; Schmader & Johns, 2003; Sekaquaptewa & Thompson, 2002). In fact, Stone & McWhinnie (2008) found that the gender of the experimenter by itself was sufficient to produce a threat effect in the context of a putting task. These
findings suggest that investigators in motor control may want to consider assessing the beliefs of their participants about gender differences in a post-experimental questionnaire to ensure that gender stereotypes have not been inadvertently activated.

In conclusion, our research demonstrated the effect of stereotype threat on sensorimotor performance supporting a motivational explanation as formulated by the mere effort account. Our experimental results show that performance outcome measures alone are not sufficient to determine the mechanisms through which stereotype threat affects task performance. When using complex sensorimotor tasks to study the effects of ST, it is important to consider how it affects the underlying control mechanism(s), even if there is no detectable effect on overall performance. By demonstrating how subtle psychological influence can differentially affect motor learning, our findings also stress the need for care in verbal instructions when conducting motor learning experiments.
6.7 Figures

Figure 6.1. Front and side view of the virtual experimental setup for ball bouncing. Participants were positioned in front of a screen and manipulated a real table tennis racket to rhythmically bounce a virtual ball to a target height in a 2D virtual environment.
Figure 6.2. Simulation of the ball-racket system. (A) Assuming sinusoidal racket movement, the racket trajectory has a segment with positive acceleration followed by negative acceleration before its peak position. (B) When the racket impacts the ball during the decelerating portion of the racket’s upward motion, the ball-racket system is dynamically stable. Slightly different initial conditions all lead to the same stable ball amplitude without any changes in the racket trajectory. (C) If the ball impacts the racket during the accelerating portion of the racket’s upward motion, the system is unstable. Different initial conditions all lead to unstable behavior where the ball finally sticks to the racket. The only way to achieve and maintain a stable pattern is to correct for errors in the ball amplitude by a change in the racket trajectory.
Figure 6.3. Schematic diagram describing the mere effort account. If the prepotent response is correct, the mere effort account suggests that stereotype threat facilitates performance for stigmatized individuals. If the prepotent response is incorrect and participants do not know, performance is debilitated. However, if participants are able to recognize that their prepotent tendencies are incorrect and have the time to correct them, performance can be facilitated.
Figure 6.4. Time series of racket (black) and ball (gray) trajectories illustrating the dependent measures. A bounce is the event between two consecutive ball-racket impacts. Error was defined as the unsigned difference between the target height and the maximum ball amplitude at each bounce. Racket acceleration at impact was defined as the racket acceleration 25 ms before the ball-racket impact of each bounce. Ball velocity at release was defined as the velocity of ball at the instantaneous ball-racket impact of each bounce. Bounces, where the difference between the ball and racket maximum positions was below 0.25 m, referred to as stuck bounces, were excluded from the data analysis in Experiments 2 and 3.
**Figure 6.5.** Performance measures of all participants over practice in Experiment 1. **(A)** The means of median of error of participants in all four experimental groups. **(B)** Means of the median racket accelerations at impact for each trial across 25 trials. **(C)** Means of lag-1 autocorrelation of release ball velocities at impact for each trial across 25 trials.
Figure 6.6. Mean error and racket acceleration across practice and statistical comparisons in Experiment 2. (A) Participant means of median errors across 25 trials. (B) Statistical comparison of error between experimental groups. Error bars represent ± 1 standard error. (C) Participant means of median racket accelerations across 25 trials. (D) Statistical comparison of racket acceleration between experimental groups.
Figure 6.7. Mean error and racket acceleration across practice and statistical comparisons in Experiment 3. Error bars represent ± 1 standard error of the mean. The Female ST group does not receive the stereotype threat manipulation until after trial 12, as marked with the vertical line. (A) Participant means of median errors across 24 trials. (B) Statistical comparison of error in Blocks 1 and 2 between experimental groups. (C) Participant means of median racket accelerations across 24 trials. (D) Statistical comparison of racket acceleration at impact in Blocks 1 and 2 between the two experimental groups.
Figure 6.8. Simulation of the ball-racket system with time-dependent manipulation. (A) If the ball impacts the racket during the accelerating portion of the racket’s upward motion, the system is unstable. (B-C) When the racket impacts the ball during the decelerating portion of the racket’s upward motion, the system is unstable if the racket phase is less than $0.05\omega$; otherwise, the perturbed ball-racket system is dynamically stable.
7. Girls *Can* Play Ball: Stereotype Threat Reduces Variability in a Motor Skill

The previous study assessed the influence of increased motivation through verbal instruction on a rhythmic performance of the ball bouncing task. This study examined the effect of the same instruction on discrete performance of the same task. Behavioral, modeling, and neuroimaging results suggest that control of rhythmic movements is distinct from the control of discrete movements (Hogan & Sternad, 2007; Howard et al., 2011; Ikegami et al., 2010; Ronsse et al., 2009; Schaal et al., 2004; Sternad, Dean, & Schaal, 2000; Sternad et al., 2013). Hence, it is likely that also the underlying mechanisms of learning are different. The previous study showed that the instruction had differential effects on performance based on the control strategy used to perform the task. This experiment aims to determine if the influence of instruction differentially affects a discrete task.

*I contributed to the experimental design, implementation of the virtual task, data analysis, statistical analysis, and interpretation of results.*

7.1 Abstract

The majority of research on stereotype threat shows what is expected: threat debilitates performance. However, facilitation is also possible, although seldom reported. This study investigated how stereotype threat influences novice females when performing the sensorimotor task of bouncing a ball to a target. We tested the predictions of two prevailing accounts for debilitation and facilitation due to stereotype effects: working memory and mere effort. Experimental results showed that variability in performance
decreased more in stigmatized females than in control females, consistent with the prediction of the mere effort account, but inconsistent with the working memory account. These findings suggest that stereotype threat effects may be predicated upon the correctness of the dominant motor behavior rather than on a novice-expert distinction or task difficulty. Further, a comprehensive understanding should incorporate the fact that stereotype threat can facilitate, as well as debilitate, performance.

7.2 Introduction

Twenty years ago, Steele and Aronson (1995) coined the term Stereotype Threat (ST) to describe the concern that arises when one feels at risk of confirming a negative stereotype about one’s group. Subsequent research has focused on how this concern debilitates the performance of stigmatized groups. For example, when examining gender stereotypes, the typical question has been, “Why do women underperform under stereotype threat?” (Cadinu, Maass, Rosabianca, & Kiesner, 2005). Consistent with the premise of this question, research has shown debilitation in a variety of cognitive and sensorimotor tasks. However, could the expectation that women underperform under ST be just another stereotype?

The focus on the debilitating effects of ST may stem from its potential negative societal implications. For example, lower ability in science and math is one of the most prominent stereotypes of females that may account for the underrepresentation of females in these fields (Eccles, Jacobs, & Harold, 1990; Nosek et al., 2009). Furthermore, S. J. Spencer et al. (1999) demonstrated that women with a strong mathematical training performed worse than men with average training on the advanced GRE exam in mathematics; they performed only equally well on a comparable GRE exam of average
difficulty. Critically, when women were told that the difficult exam did not produce gender differences, they performed as well as men, suggesting that stereotypes about math ability had influenced their performance. Thus, a better understanding of ST effects may prevent the failure of young women and encourage and enable them to pursue careers in STEM fields – e.g., 28% of STEM tenure-track faculty in the US were female in 2013 (National Science Foundation, 2013).

In fact, the current accounts of ST effects flow directly from this emphasis on debilitation in the cognitive domain. The prevailing perspective argues that concern over confirming the stereotype produces disrupting thoughts that utilize cognitive resources, which could be otherwise devoted to task performance. It is this reduction in working memory capacity that causes the debilitation so often reported on cognitive tasks (Schmader, Hall, & Croft, 2015; Schmader et al., 2008).

The effects of ST on motor performance have also been studied, although much less extensively than on cognitive tasks. Again, most research regarding the effect of ST on sensorimotor performance has observed debilitating effects. Studies have reported debilitation from ST in a variety of sensorimotor tasks such as golf putting (Beilock et al., 2006; Stone et al., 1999; Stone & McWhinnie, 2008), soccer dribbling (Chalabaev et al., 2008; Heidrich & Chiviacowsky, 2015), simulated driving (Yeung & von Hippel, 2008), tennis serving (Hively & El-Alayli, 2014), and basketball free throw shooting (Hively & El-Alayli, 2014; Krendl, Gainsburg, & Ambady, 2012). The majority of these studies examined the effects of a gender-related ST, reporting that female performance is debilitated when exposed to the stereotype that females perform worse than males either in athletic performance or in that specific motor task (Chalabaev et al., 2008; Heidrich &
While less commonly studied, it has also been shown that male performance in golf putting can be debilitated when instructed that females perform this task better than males (Beilock et al., 2006). In addition, evoking race-related stereotypes has led to debilitated sensorimotor performance in the stigmatized group (Krendl et al., 2012; Stone et al., 1999). These reports are consistent with the pervasive stereotype that males are more competent in athletics (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002) and show higher levels of daily physical activity (Knisel, Opitz, Wossmann, & Keteihuf, 2009).

To explain the debilitating effects in motor performance, Schmader et al. (2015) suggest that ST increases performance monitoring, which in turn reduces working memory and disrupts task execution. This is particularly noticeable in well-learned, proceduralized tasks such as golf putting. However, a recent study by Huber, Seitchik, et al. (2015) found that the same ST manipulation could be used to debilitate and facilitate motor performance under different circumstances. While any observation of facilitated performance under ST is incongruent with the predictions of the working memory account, the findings of Huber, Seitchik, et al. (2015) were consistent with an alternative account developed by Jamieson and Harkins (2007). This account, referred to as “mere effort,” argues that individuals faced with ST are motivated to disprove the negative stereotype about their group, leading to the potentiation of the dominant, or prepotent, response. For sensorimotor tasks, the prepotent response is considered the dominant motor behavior, which can either be correct or incorrect, depending on whether or not the dominant motor behavior leads to the desired performance of the task. Huber, Seitchik, et
al. (2015) reported that ST affected performance in a rhythmic ball bouncing task in opposite ways, depending on the correctness of the prepotent response. This response was determined by the skill level of the performer: In novices, the prepotent response was incorrect, and therefore ST debilitated their performance; for those experienced in the task, the prepotent response was correct, and ST therefore facilitated their performance. This latter finding highlighted a largely neglected fact: under certain conditions, women may actually rise to the challenge and improve their performance under ST (Jamieson & Harkins, 2007, 2009, 2011; O’Brien & Crandall, 2003). We believe that research on facilitation under ST is very relevant, since a better understanding of how and when ST facilitates performance can also help us better understand conditions under which ST debilitates performance.

In Huber, Seitchik, et al. (2015), facilitation due to ST was only observed for performers experienced with the task. All prior work investigating the effect of ST on motor performance for novices has reported debilitation (Heidrich & Chiviacowsky, 2015; Krendl et al., 2012; Stone & McWhinnie, 2008). In the current work, we asked if ST could also facilitate the performance of inexperienced performers on a novel sensorimotor task. Following our previous results that it is the dominant behavior that determines the effect of ST, we chose a task where this dominant behavior was correct. Unlike the distinction between novices and experts, the mere effort account grounds its predictions on the correctness of the dominant or prepotent behavior. Thus, in order to observe facilitation in novices, we first had to identify a motor task where the dominant behavior was correct in novice subjects. Given the correct dominant response, the mere
effort account predicted that novice performance would be facilitated. In contrast, the working memory account predicted debilitation.

7.3 Baseline Experiment

The purpose of the baseline experiment was to identify a task where the dominant behavior of novices was correct and quantify this behavior. The experiment introduced a discrete version of the ball bouncing task, where subjects hit a ball to a target line in a single bounce. This task resembled the golf putting accuracy task frequently used in prior ST research (Beilock et al., 2006; Stone et al., 1999; Stone & McWhinnie, 2008). In aiming tasks, errors in motor performance can be caused by a constant bias (e.g., tendency to under- or overshoot the target) and/or by variability around the desired solution (Schmidt & Lee, 2011). A constant offset would suggest that the prepotent response was incorrect, whereas the absence of a bias (i.e. variability is clustered evenly around the target) would suggest that the prepotent response was correct.

It is important to note that while the experimental setup of the discrete ball bouncing task was similar to the rhythmic ball bouncing task used in our previous experiments (de Rugy et al., 2003; Dijkstra et al., 2004; Ehrlenspiel et al., 2010; Huber, Seitchik, et al., 2015), the motor control demands were very different as different motor strategies are used in discrete versus continuous rhythmic performance (Hogan & Sternad, 2007).

7.3.1 Methods

7.3.1.1 Participants

25 undergraduate students (13 males and 12 females) from Northeastern University participated in the experiment in exchange for partial fulfillment of a course requirement.
None had any prior experience with the specific task. Prior to the experiment participants read and signed the consent form as approved by the Institutional Review Board of Northeastern University. We planned to recruit an equal number of males and females, however data collection had to be terminated at the end of the semester, leading to the slightly uneven numbers.

7.3.1.2 Task

In the experimental task, the participants used a real racket to bounce a virtual ball to a target line (for a detailed description of the experimental setup, see K. Wei et al., 2007). The participants stood in front of a projection screen holding a real table tennis racket in his or her dominant hand (Figure 7.1). The screen displayed a virtual scene consisting of a ball, a racket, a target line positioned 1.0m above the racket, and a number score. The vertical displacements of the real racket controlled the vertical position of the virtual racket.

At the start of each trial, the ball appeared at the left side of the screen and then rolled horizontally along the target line to the center of the screen (Figure 7.1). Upon reaching the center, the ball dropped vertically from the target line towards the virtual racket. The participant was instructed to bounce the ball such that the maximum ball height was within ±3cm of the center of target line. For ball amplitudes higher or lower than this distance, the ball no longer overlapped with the target line. The vertical position of the virtual ball after ball-racket impact was determined using the equations for ballistic flight (see Appendix C). Successful bounces within ±3cm of the target line were signaled with a temporary color change of the target line, which acted as a reward signal to the subjects. Participants were instructed to produce as many successful bounces as possible. The number of successful bounces in each block was displayed as a score on the top right
corner of the screen. Following a brief pause, the next trial began. Each trial or bounce lasted approximately 3.5 secs. All participants were given two practice trials for familiarization and then completed 12 blocks of 30 trials each under the watch of the experimenter, with a short break after block 6. The experiment lasted approximately 30 minutes.

7.3.1.3 Dependent Measures

The first measure to characterize task performance was the percentage of successful bounces in each block of 30 trials/bounces. Bounces were deemed successful when the error was between ±3 cm. The second measure was error, defined as the signed difference between the maximum ball height and the target line (1.0 m). The median of errors of the 30 trials/bounces in a block was calculated to serve as a measure of central tendency. Median error values outside the success region (±3 cm) would indicate that subjects had a systematic bias or a tendency to under- or overshoot the target line. Variability, the third measure, was quantified by the interquartile range (IQR) of error in each block. IQR, the range of the second and third quartile of the distribution, is a frequently used measure to estimate dispersion. Median and IQR were used as Shapiro-Wilk tests revealed that the distributions of error were not normal in approximately 3 out of 12 blocks for each subject (Shapiro & Wilk, 1965).

7.3.1.4 Statistical Analyses

The three dependent measures were analyzed with a 2 (Gender) x 12 (Block) ANOVA, with gender as between-subject factors, and block as a within-subject factor. The Greenhouse-Geisser correction factor was applied to the within-subject effects (Kirk, 1995). Relevant planned comparisons using independent sample t-tests investigated group effects.
7.3.2 Results

7.3.2.1 Task Success

The percentage of successful bounces per block was used to measure overall task performance and its improvement with practice (Figure 7.2A). The ANOVA revealed a main effect for block, $F(11, 253) = 9.72$, $p < .001$, $\eta^2_p = .30$, indicating that the percentage of successful bounces increased over the course of the 12 blocks. The ANOVA also yielded a weakly significant Gender x Block interaction, $F(11, 253) = 2.14$, $p = .042$, $\eta^2_p = .085$. Planned comparisons revealed that there were no significant differences between males and females on any blocks except the last block where males performed significantly better ($M=33.33\%, SD=11.94\%$) compared to females ($M=23.89\%, SD=10.72\%), t(23) = 2.07, p=.049. There was no significant main effect for Gender, $F(1, 23) = .59, p > .250$. While participants did increase their percentage of successful bounces with practice, they never exceeded more than 50% success (Figure 7.2A). This signaled that participants were indeed novices and that the task was challenging. Table 7.1 presents the overall means and standard deviations for the dependent measures for each gender.

7.3.2.2 Median Error

In contrast to the successful bounces, the median error showed no significant change across blocks as confirmed by the ANOVA, $F(11, 253) = 1.79, p = .162$ (Figure 7.2B). All participants showed median errors that were in the gray success region, implying that they accurately centered their performance on the target line. There was no significant effect of gender, $F(1, 23) = .82, p > .250$, nor an interaction, $F(1, 23) = .56, p > .250$ (see Table 7.1). This observation indicated that subjects did not have a constant
bias in their motor performance. Thus, the dominant behavior, or prepotent response of novices, was correct, both in males and females.

7.3.2.3 Variability of Error

The variability measure, IQR, quantified the distribution of errors and expressed how precisely participants hit the target line. As for the number of successful bounces, a main effect of block was revealed by the ANOVA, $F(11, 253) = 19.43, p < .001, \eta^2_p = .46$ (Figure 7.2C). All participants decreased their variability with practice. There was no significant effect of gender, $F(1, 23) = .96, p > .250$, nor an interaction, $F(1, 23) = 2.00, p = .132$ (Table 7.1). It should be noted that unlike in the measure of task success, a gender difference did not emerge in this measure, even though it is more fine-grained and has higher-resolution.

7.3.3 Discussion

Results showed that novices accurately centered their performance on the target line from the outset of the performance. As there was no constant offset, it was concluded that the dominant response was correct, even though the participants were all novices to the task. Performance improvements were the outcome of refining this dominant behavior by decreasing their variability and thereby increasing the percentage of successful bounces.

7.4 Experiment with Stereotype Threat

Having established the predominant behavior in the discrete bouncing task, this experiment introduced a stereotype threat manipulation. This instruction, shown to be effective in prior work, implied that females would show inferior performance in this visual-spatial task and that it was related to math ability. Importantly, extending from
results in the baseline experiment, differential predictions could be made for the working memory account and the mere effort account. The mere effort account predicts that ST would potentiate the overall correct behavior and thereby improve or facilitate the performance of novices. On the other hand, the working memory account hypothesizes that novices should show inferior performance under ST, although only if the novel task “tests the upper bound of one’s skill level” (Schmader et al., 2015, p. 450). Given that subjects in the baseline experiment still performed with low success rates by the end of practice, the working memory account predicts that novices under ST show debilitated performance.

7.4.1 Methods

7.4.1.1 Participants

48 undergraduate students (24 males and 24 females) from Northeastern University participated in the experiment in exchange for partial fulfillment of a course requirement. None participated in the baseline experiment or had any prior experience with the specific task. Prior to the experiment participants read and signed the consent form as approved by the Institutional Review Board of Northeastern University.

7.4.1.2 Task and ST Manipulation

The task and procedure were identical to that of the baseline experiment, but with two additions. First, after completing the two practice bounces prior to the first block of trials, a male experimenter administered the following verbal instructions to participants in the ST condition:

The task you are about to complete is a test of visuo-spatial capacity. Performance on this task is closely linked to math ability. As you may know, there has been some controversy about whether there are gender differences in math and spatial
ability. Previous research has demonstrated that gender differences exist on some of these tasks, but not on others. In our lab, we examine performance on both kinds of tasks. The task on which you are about to participate has been shown to produce gender differences.

The instruction implied that males outperform females in this task, which is a negative stereotype for females. For males this instruction presented a negative stereotype about another group. Therefore, only the females who received the manipulation experienced stereotype threat. This verbal instruction has been shown to produce ST effects in previous research (e.g., Brown & Pinel, 2003; Keller & Dauenheimer, 2003; O’Brien & Crandall, 2003; Spencer et al., 1999). Participants in the No-Threat (NT) condition heard the same manipulation with the sole difference that the final sentence read: “The task on which you are about to participate has not been shown to produce gender differences.”

Second, upon completion of block 12, participants filled out 11-point scales that assessed the effectiveness of the ST manipulation. The effect of the ST manipulation was gauged based on responses to two questions: (1) “To what extent do you believe that gender differences exist on this task?” (1 indicated no gender difference and 11 reflected gender differences); and (2) “Who do you believe performs better on this task?” The scale’s midpoint at 6 indicated that males and females performed the same, while lower values indicated males performed better than females and higher scores reflected the converse.

7.4.1.3 Statistical Analyses

The same dependent measures as in the baseline experiment were analyzed in 2 (Gender) x 2 (Threat) x 12 (Block) ANOVAs, with gender and threat as between-subject
factors, and block as a within-subject factor. The Greenhouse-Geisser correction factor was applied to the within-subject effects (Kirk, 1995). The analyses of the manipulation check excluded the block factor, but were otherwise identical.

7.4.2 Results

7.4.2.1 Manipulation Check

A 2 (Gender) x 2 (Threat) between-subjects ANOVA indicated that participants in the ST condition believed that gender differences existed to a greater extent ($M = 6.71$, $SD = 2.37$) than the participants in the NT condition ($M = 2.67$, $SD = 2.04$), $F(1, 44) = 38.46$, $p < .001$, $\eta^2_p = .47$. Neither the gender main effect, nor the interaction was significant, $ps > .250$. Participants in the ST condition also reported that males outperformed females on this task to a greater extent ($M = 3.71$, $SD = 1.08$) than NT participants ($M = 5.38$, $SD = .92$), $F(1, 44) = 31.54$, $p < .001$, $\eta^2_p = .42$. Again, neither the gender main effect, nor the interaction was significant, $ps > .250$. These results indicated that the manipulation not only successfully induced ST in females, but that the ST instruction also conveyed a positive stereotype for males.

To further confirm that there was no influence of the NT instruction on behavior, the behavior of participants in the NT condition was compared to that of the participants in the baseline condition. All dependent measures were analyzed using 2 (Gender) x 2 (Condition) x 12 (Block) ANOVAs, with gender and condition (NT or baseline) as between-subject factors, and block as a within-subject factor. The analyses revealed no significant main effects for condition or gender, nor any significant interactions on any measure (see Appendix D).
7.4.2.2 Task Success

The percentage of successful bounces per block was used to measure overall task performance (Figure 7.3). The ANOVA revealed a main effect for block, $F(11, 484) = 46.27, p < .001, \eta_p^2 = .51$, indicating that the percentage of successful bounces increased over the course of the 12 blocks. Replicating the results of the baseline experiment, subjects in this experiment never exceeded more than 50% success (Figure 7.3). The analysis also revealed a main effect for threat, $F(1,44) = 14.57, p < .001, \eta_p^2 = .25$, showing that all participants in the ST condition had a higher percentage of successful bounces ($M = 30.8\%, SD = 8.0\%$) than their non-threatened counterparts ($M = 21.5\%, SD = 9.3\%$). While both males and females in the ST condition showed better performance, it is important to note that the females acted in response to a negative stereotype to themselves, whereas males performed better upon hearing a negative stereotype about others.

The ANOVA also yielded a significant Threat x Block interaction, $F(11, 484) = 2.43, p = .017, \eta_p^2 = .05$. While there was no significant difference between the conditions in block 1, group differences emerged over the course of the task. The main effect for gender did not reach significance, $F(1,44) = 2.72, p = .106$, nor any of the remaining interactions, Gender x Block, $F(11, 484) = 1.17, p > .250$, Gender x Threat, $F(1, 44) = .26, p > .250$, and Gender x Threat x Block $F(11, 484) = .58, p > .250$. Table 7.2 presents the overall means and SDs for each condition and gender.

7.4.2.3 Median and IQR of Error

Figure 7.4 illustrated how the distribution of error changes over blocks in two example female participants, one in the ST condition and one in the NT condition. Initially, performance was highly dispersed, but nevertheless clustered over the success
region highlighted in gray for both participants. As summarized in Figure 7.5A, this pattern was representative for all participants, implying that they accurately centered their performance on the target line right from the onset of practice. In accordance with this observation, the median error showed no significant change across blocks or conditions as confirmed by the ANOVA ($p > .250$; Table 7.2). This observation indicated that the prepotent response of novices, seen already in the baseline experiment, was correct for this task and did not change with practice.

Submitting the IQR of error to the 2 (Gender) x 2 (Threat) x 12 (Block) ANOVA revealed a main effect for threat, $F(1, 44) = 17.70, p < .001, \eta^2_p = .29$. Participants in the ST condition hit the ball with less variability or smaller IQR ($M = 12.6cm, SD = 3.6cm$) than the NT participants ($M = 20.2cm, SD = 8.0cm$). This analysis also revealed a significant main effect for block, $F(11, 484) = 46.27, p < .001, \eta^2_p = .51$. A significant Threat x Block interaction, $F(11, 484) = 2.85, p = .021, \eta^2_p = .06$, resulted from the fact that variability dropped significantly from block 1 to block 2, $F(1, 484) = 12.10, p < .001$, for the ST participants, but not for NT participants, $F < 1$. There was neither a main effect for gender, $F(1,44) = 1.58, p = .216$, nor interactions of Gender x Block, $F(11, 484) = 1.17, p > .250$, Gender x Stereotype, $F(1, 44) = .26, p > .250$, and Gender x Stereotype x Block, $F(11, 484) = .21, p > .250$ (see Table 7.2).

To assess the robustness of the results in this experiment, we also compared the behavior of participants in the ST condition to the participants in the baseline condition (see Appendix E). Overall, the results consistently demonstrated that the ST enhanced performance. While the results of this analysis revealed similar effects of ST, caution
needs to be applied as we did not control for any stereotype that participants in the baseline experiment may have held.

7.4.3 Discussion

As in the baseline experiment, the median error was in the success region right from the outset and was not affected by the threat instruction. In contrast, the variability measure improved in all participants under the ST instruction, leading to a higher percentage of successful bounces. This facilitation, induced by the ST instruction, is consistent with the motivational mere effort account and inconsistent with the working memory account. Based on the overall correct behavior observed in the baseline experiment and corroborated in the ST experiment, the mere effort account argues that participants should be motivated by ST and perform better through potentiation of the prepotent response.

We also found that males in the ST condition outperformed their male NT counterparts. To understand this finding, it should be kept in mind that the same verbal instruction was probably perceived differently by each gender. Thus, the facilitated performance of males in the ST condition did not result from stereotype threat, but rather may have resulted from stereotype lift. We discuss this distinction further in the general discussion.

7.5 General Discussion

Stereotype threat research has primarily focused on, and repeatedly produced, debilitated performance on cognitive and motor tasks. While this addresses a serious societal concern, we argue that under certain conditions ST may also improve performance. Therefore, any account of ST effects must give serious consideration to the
possibility that ST can both facilitate and debilitate performance. With this goal, the current study examined conditions where ST could indeed facilitate performance, particularly in novices.

7.5.1 **Eliciting Threat by Evoking a Negative Stereotype**

The verbal instruction implied that males should outperform females in this discrete ball bouncing task. This same verbal instruction has been shown to produce ST effects in several previous studies (e.g., Brown & Pinel, 2003; Keller & Dauenheimer, 2003; O’Brien & Crandall, 2003; Spencer et al., 1999). Huber, Seitchik, et al. (2015) also reported that this instruction could produce both debilitation and facilitation in stigmatized females. Thus, it is safe to conclude that this instruction contains no bias that may elicit facilitation. As subjects never reached a high level of performance, it is also unlikely that the facilitation was due to a “challenge” response to the stereotype instruction. Such a response occurs when subjects believe their abilities are sufficient to meet the task demands. It differs from a “threat” response, where subjects assess their abilities as insufficient for accomplishing the task (Blascovich & Mendes, 2000). Based on these considerations, we posit that the facilitated performance of stigmatized females was in fact due to stereotype threat.

7.5.2 **Effect of Stereotype Threat on Motor Variability**

Before discussing the theoretical implications of the findings from the ST experiment, it is important to highlight that this facilitating effect was evidenced by a reduced variability in the ST group, not by a change in the central tendency (i.e. mean or median) as typically considered. In basic motor control research, variability in task performance has been widely recognized as an essential window into control processes (Davids, Bennett, & Newell, 2006; Newell & Corcos, 1993; Sternad et al., 2011).
Variability is a measure independent from average behavior and analyses of its temporal and distributional structure have revealed numerous insights into sensorimotor control (Abe & Sternad, 2013; Cohen & Sternad, 2009; Gilden, Thornton, & Mallon, 1995; Sternad, Dean, & Newell, 2000). Unlike in cognitive tasks, where the verbal or written output directly reflects the response from executive centers, motor performance is subject to many additional peripheral and central processes that add variability to the initial plan. Hence, measures of variability present informative variables that have not yet been exploited in studies of ST on performance.

### 7.5.3 Theoretical Implications of Experimental Results

The goal of the study was to compare the predictions of the working memory and mere effort accounts on sensorimotor performance of female novices under ST. The working memory account predicted debilitation for novices in a challenging task. In contrast, the mere effort account predicted facilitation, if the predominant behavior was regarded correct. Creating a task scenario that was challenging but nevertheless showed a correct behavior even in novices, participants under ST were shown to improve their performance by decreasing their variability and increasing their successful target hits.

Schmader et al. (2015) suggest that ST increases performance monitoring, which in turn reduces working memory and disrupts task execution on well-learned, proceduralized tasks. However another account, the explicit monitoring account, proposes different underlying processes. The explicit monitoring account suggests that it is the availability of working memory, rather than its absence, that produces this debilitation on well-learned tasks (Beilock et al., 2006). The authors propose that ST causes experts to attend closely to sensorimotor processes and this explicit attention to step-by-step processing is thought to disrupt task execution that normally runs outside of
conscious awareness (Baumeister, 1984; Beilock, Bertenthal, McCoy, & Carr, 2004; Gray, 2004; Langer & Imber, 1979; Masters, 1992). If ST prompts explicit monitoring as suggested, it then follows that for novices ST-induced explicit monitoring should facilitate the performance (Beilock et al., 2004; Beilock, Carr, MacMahon, & Starkes, 2002; DeCaro et al., 2011; Gray, 2004). The results of this experiment support these predictions. However, the explicit monitoring account does not explain the debilitation that is in fact more commonly observed in stigmatized novices. Recently, Chalabaev and colleagues also questioned both working memory and explicit monitoring accounts, as they observed that ST debilitated performance on a ballistic isometric force task, even though the task was too short to allow any explicit monitoring processes relying on working memory (Chalabaev, Brisswalter, et al., 2013).

The mere effort account explains the present findings and also other experimental results, previously reported as support for the explicit monitoring account. For example, (Beilock et al., 2006) showed that ST debilitated expert players in a golf putting task and attributed this performance decrement to undue attention to processes that usually run automatically. While plausible, their findings can also be explained by the mere effort account. In their task, golf experts were asked to putt the ball so that it stopped directly on the target. This requirement differed from how these experts typically putt, making the task more difficult. Putting the ball through the target and overshooting is the most typical strategy for expert golfers, which was the incorrect behavior for this experiment. Hence, the reported debilitation would be consistent with the mere effort account.

7.5.4 Novice and Expert Performance Versus Correct and Incorrect Behavior

Unlike the other accounts, mere effort account does not distinguish between novices and experts, rather, between correct and incorrect prepotent behavior. In the
rhythmic ball bouncing task and in a discrete golf putting task, ST debilitated novice performance (Huber, Seitchik, et al., 2015; Stone & McWhinnie, 2008), whereas ST facilitated the performance of stigmatized novices in this discrete ball bouncing task. These seemingly contradictory results raise the possibility that the distinction between “novice” and “expert” may not be the most appropriate to explain ST effect on sensorimotor performance (Chalabaev et al., 2008; Stone & McWhinnie, 2008). The notion that performance under threat is contingent upon the correctness of the prepotent response is consistent with other accounts that suggested that the effect of ST depends on whether the task was “easy” or “hard” (O’Brien & Crandall, 2003) or whether the task is “well-learned” or not (Beilock et al., 2006). While tasks characterized as “easy” often have correct prepotent responses, this is not always the case. Harkins (2006) showed that prepotency precedes task difficulty, when predicting performance outcomes for stigmatized individuals. The same is true of the relationship between task experience and prepotency (i.e., experts do not always have a correct prepotent response). Furthermore, there is no consensus or quantitative criteria for such categorizations. Evaluating the behavior, rather than the task offers a more principled criterion, as well as more effective means to making predictions for performance under threat.

7.5.5 Stereotype Lift in Males

A secondary finding was that the male subjects who received the same verbal instruction also outperformed their NT counterparts. This behavior is consistent with the notion of stereotype lift, previously observed in males in ST research (Chalabaev, Sarrazin, et al., 2013; Chatard, Selimbegović, Konan, & Mugny, 2008; Laurin, 2013; Shih, Ambady, Richeson, Fujita, & Gray, 2002; Shih, Pittinsky, & Ambady, 1999; Walton & Cohen, 2003; Wraga, Helt, Jacobs, & Sullivan, 2007). Stereotype lift has been
suggested to increase motivation to uphold the manipulated stereotype and buttress self-esteem. While the behavioral results show similar decreases in variability for both genders in the ST condition, these results are likely to have occurred through different processes. The same verbal instruction conveyed very different messages to the participants: for example, males may have been motivated to confirm the negative stereotype about females, and by extension, confirm a positive stereotype about males. In contrast, females are motivated to disconfirm the negative stereotype. Indeed, other research on sensorimotor performance, activating negative stereotypes about females led males to perform better through increases in effort (Chalabaev et al., 2008).

7.5.6 Study Limitations

While the experimental results provide support for the mere effort account, the study also had also some limitations. A first caveat is the identification of the prepotent response. While we could identify the prepotent response through behavioral data, it is not possible to extract the prepotent response at the neural level. Second, this task could not test the mere effort account’s prediction of debilitation. However, in the previous study that used the rhythmic version of this task (Huber, Seitchik, et al., 2015), both facilitation and debilitation was predicted and observed. Note, though, that performance measures of rhythmic and discrete execution are different. Hence, an interesting experimental test would be to use the discrete task variant and teach subjects the wrong prepotent response to then test both predictions.

One additional limitation was that we only examined the effect of ST on the initial performance of novices. It is just as important to know whether the effect of ST would persist if novices continued to improve task performance over several days of practice.
Nevertheless, the results of this study clearly mark the need for a more comprehensive explanation of ST including facilitation under several different circumstances.

7.6 Conclusions and Outlook

While a great deal of research has focused on explaining how ST debilitates performance, any comprehensive account should also be able to explain how ST can facilitate performance. Not only do we observe facilitation under ST in the lab, but we also see women surpass men at elite sports despite the still prevailing female stereotype in athletics. Famous examples are boxer Jackie Tonawanda and the bowler Kelly Kulick. Perhaps by understanding the conditions in which ST improves performance, we may gain insights on how to help females perform better in the presence of pervasive societal stereotypes. The results of this study also underscored that measures of motor variability are essential to understand how subtle psychological manipulations affect performance in a sensorimotor task. Hence, we agree with the editors of a recent handbook on ST who referred to their volume as only the “halftime report” (Michael Inzlicht & Schmader, 2012) and argue that a more comprehensive account of ST is required.
7.7 Figures and Tables

Figure 7.1. Side and front view of the virtual experimental setup for discrete ball bouncing. Participants were positioned in front of a screen and manipulated a real table tennis racket to bounce a virtual ball to a target height in a 2D virtual environment.
Figure 7.2. Results of baseline experiment. (A) Task success, (B) median error, and (C) IQR error over blocks. Each point represents the group average of the dependent measures per block, and the error bar represents the standard error across participants in each group.
Figure 7.3. Task success in ST experiment. Each point represents the group average of task success per block, and the error bar represents the standard error across participants in each group.
Figure 7.4. Distribution of errors in each block for two example female participants from the NT group and the ST group. Each circle represents one trial or bounce. The shaded region represents errors small enough to be deemed a successful bounce (±3cm).
Figure 7.5. (A) Median and (B) IQR of error in ST experiment. Each point represents the group average of the dependent measures per block, and the error bar represents the standard error across participants in each group.
Table 7.1. Overall means and standard deviations of each dependent measure for the between-subject factor in the Baseline Experiment.

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M (SD) )</td>
<td>( M (SD) )</td>
</tr>
<tr>
<td>% of Successful Bounces</td>
<td>23.36 (7.52)</td>
<td>21.27 (5.85)</td>
</tr>
<tr>
<td>Median Error (cm)</td>
<td>0.63 (1.80)</td>
<td>1.25 (1.62)</td>
</tr>
<tr>
<td>IQR of Error (cm)</td>
<td>16.73 (5.91)</td>
<td>19.03 (5.86)</td>
</tr>
</tbody>
</table>

Table 7.2. Overall means and standard deviations of each dependent measure for between-subject factors in the Stereotype Threat Experiment.

<table>
<thead>
<tr>
<th></th>
<th>Males NT</th>
<th>Males ST</th>
<th>Females NT</th>
<th>Females ST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M (SD) )</td>
<td>( M (SD) )</td>
<td>( M (SD) )</td>
<td>( M (SD) )</td>
</tr>
<tr>
<td>% of Successful Bounces</td>
<td>24.86 (11.31)</td>
<td>31.51 (8.08)</td>
<td>18.15 (5.13)</td>
<td>30.16 (8.19)</td>
</tr>
<tr>
<td>Median Error (cm)</td>
<td>1.58 (2.45)</td>
<td>0.93 (2.31)</td>
<td>0.86 (2.94)</td>
<td>0.22 (2.43)</td>
</tr>
<tr>
<td>IQR of Error (cm)</td>
<td>18.60 (9.03)</td>
<td>11.97 (3.48)</td>
<td>21.77 (6.88)</td>
<td>13.31 (3.66)</td>
</tr>
</tbody>
</table>
8. Discussion and Future Directions

The goal of this thesis was to better understand the mechanisms of complex skill learning in order to inform future advancements in motor rehabilitation, in particular using VR and robotic rehabilitation. From a patient’s perspective, rehabilitation involves relearning how to move in order to regain functional independence in daily life. Therefore our knowledge of motor control and learning should expand to reveal how people learn functionally relevant and complex tasks. However, conducting the breadth of research required for a comprehensive understanding of complex skill learning is no small feat, especially with the added objective of understanding the influence of neural impairment on complex skill. Hence, the focus of the questions in this thesis had to be narrowed.

Specifically this research concentrated on understanding complex skill learning from a behavioral perspective and how interventions should be designed to enhance learning by using virtual environments. The results of this research yielded several novel insights into complex skill learning and interventions:

1. Virtual tasks present an ideal and versatile platform for theoretically-grounded and fine-grained quantitative assessment of motor behavior during complex skill learning.

2. Strategies used to learn simple tasks are different than those used to learn complex tasks with redundancy due to the possibility of noise-tolerant solutions in complex tasks.

3. Guidance approaches for complex skills should be distinct from those used to enhance simple task learning due to differences in the solution space.
Irrespective of the task, guidance methods should be designed to promote persistence of the learned behavior upon removal of the assistance.

4. Increasing motivation through verbal instruction has differential effects on task performance depending on the control strategy and the type of task.

These four advances and their implications for the development of future rehabilitation approaches are further detailed and discussed in the sections below.

8.1 Virtual and Robotic Technologies for Quantifying Motor Skill Performance

This thesis outlined a general approach that capitalizes on virtual technologies for understanding skill learning. This detailed description of the approach provides a blueprint along with methodological considerations for developing future experimental paradigms to study complex skill learning. One advantage of this research is that its model-based approach afforded development and application of many fine-grained measures to assess motor behavior during complex skill learning. These measures allowed us to examine how a range of experimental conditions affected subtle aspects of control in a way that could not be captured with gross descriptive outcome measures of motor performance such as task error and its variability. It is important to emphasize that these theoretically-grounded metrics of control were afforded by the virtual rendering of real-world tasks.

Rendering a real-world task in a virtual environment requires that the task first be mathematical modeled. Not only does the model allow for dynamic interaction with the virtual task, it also simplifies the real phenomenon by formalizing the assumptions and
removing the irrelevant aspects present in the real-life task. As a result, more direct and informative measures of control and learning over practice can be extracted.

Analysis of the mathematical model also allows one to derive hypotheses for how a given task is learned. This was demonstrated in the studies using the ball bouncing and throwing tasks. Based on the mathematical model of the task, the mapping between human behavior and task performance could be derived. This mapping was analyzed to reveal additional properties of solutions, such as risk or dynamic stability. Depending on the task model, different mathematical tools can be used to derive predictions about stability or noise-tolerance and interventions to guide learning desirable solutions.

In order to test these hypotheses, the control measures must be accurate, which means that first, human behavior must be accurately measured. Exact measures of human behavior are also critical for simulating interaction with the virtual task. Large measurement errors or delays can lead to erroneous feedback, which can ultimately hinder learning (Schmidt & Lee, 2011). Hence, accurate sensors with high spatiotemporal resolution are needed to capture human motion during performance on complex skills.

The need for high accuracy sensors is similarly critical for quantifying complex behavior during rehabilitation. Virtual rehabilitation systems utilizing the Kinect have already been used in several feasibility studies, including a rehabilitation study for children with cerebral palsy (Y.-J. Chang, Han, & Tsai, 2013) and another for patients with Parkinson’s Disease (Pompeu et al., 2014). Although the Kinect can capture human motion data with short latency, two studies included in Appendices F and G of this thesis found large inaccuracies in the measurements compared to the gold and clinical standards. Even though these studies assessed the older Kinect for Windows sensor, more
recent studies reveal that the same problems still plague the newer Kinect for Xbox One sensor (Clark et al., 2015; Xu & McGorry, 2015). These results highlight that developing new sensor and methods for measuring motor behavior, especially at a low cost, is still high priority for rehabilitation practice, especially as the behavior becomes increasing complex.

8.2 Differences Between Simple and Complex Motor Skill Learning

Prior work of Sternad and colleagues has demonstrated support for the tenet that in complex tasks with redundancy, learners “seek solutions where their noise matters less”. To further test this hypothesis, we examined how healthy subjects learned to perform a virtual throwing task over 11 days of practice. Counter to previous tenets that subjects with lower variability in timing of ball release showed better performance (Hore et al., 1996), we found that after the first practice session, there was no correlation between timing variability and task performance. Instead, our results indicated that subjects developed arm trajectories that were insensitive to timing errors and, consequently, showed better task performance. This result is consistent with the above-stated hypothesis that humans seek noise-tolerant solutions.

This research contributes a possible explanation for the discrepancy in results, suggesting that differences in solution space of the tasks may lead to differences in controlled variables. Every task has a distinctly shaped success manifold and subjects may not always be able to align their arm trajectory due to biomechanical and other physical constraints. In addition, not all tasks have redundancy, i.e. multiple solutions that achieve a given error value. For tasks where subjects cannot align their arm trajectory along the success manifold due to physical constraints or lack of task
redundancy, reducing variability is the primary strategy for improving performance. To demonstrate this point, a second experiment was conducted where subjects practiced a version of same virtual throwing task where it was not possible to develop a timing-tolerant arm trajectory. For this version of the task, results indicated that subjects’ ability to reduce variability in timing determined their overall task performance.

The novel results from these two experiments suggest that humans are sensitive to the redundancy of a task and utilize different approaches for learning based on the whether or not the redundancy offers “noise-tolerant” solutions. This insight has significance for the development of therapeutic interventions. For instance, it supports that principles for learning simple tasks do not guarantee to translate into useful principles for complex skill learning (Wulf & Shea, 2002).

8.2.1 Guiding Behavior During Motor Skill Learning

As learning objectives depends on the type of redundancy of the task, the guidance methods used for shaping behavior should be tailored to this redundancy. Having shown the significant role of task redundancy in the previous study, a subsequent study developed a method for expediting learning of the version of throwing task where reducing variability was crucial for task performance. The first novel result of this study was the demonstration of how variability naturally decreased over long-term practice, without any explicit assistance or specific feedback. We verified that subjects reduced variability predominantly by attenuating neuromotor noise, and not by increasing error correction as typically reported and modeled (Herzfeld et al., 2014; Smith et al., 2006; Thoroughman & Shadmehr, 2000). Additional experimental results and computational modeling results revealed that increasing the challenge for reward could guide subjects to reduce their neuromotor noise and improve their performance even further.
While the understanding complex skill learning is crucial for motor rehabilitation, this should not deny that the training of simple movements is also beneficial. For instance, it has been shown that training individual joint motion is more effective than training whole arm reaching movements in stroke (Krebs et al., 2008; Milot et al., 2013; Platz et al., 2005). Thus, developing interventions for simpler tasks where reducing variability may be the only route to improve task performance is still important. However, given that complex skill acquisition is the ultimate goal of motor rehabilitation, the training protocols even of simple movements should designed with this ultimate objective in mind (Schweighofer, Choi, Winstein, & Gordon, 2012).

Another approach to facilitate performance of non-redundant tasks involves manipulating variability. For instance, amplifying error has been shown to improve adaptation of reaching in healthy subjects and stroke patients (Milot et al., 2010; Patton et al., 2006; Reinkensmeyer & Patton, 2009; Sharp et al., 2011; Y. Wei et al., 2005). Amplifying variability and adding noise has also been shown to facilitate the learning of non-redundant versions of a virtual throwing task in healthy subjects (Hasson et al., n.d.) and children with dystonia (Chu et al., 2013). In a redundant line reaching task, however, adding noise did not help subjects find the most robust solutions as initially expected (Manley et al., 2014). Therefore, interventions designed solely to increase precision may not always enhance complex skill learning.

As demonstrated in this thesis, lower variability and noise do not necessarily correlate with better performance in redundant tasks. Instead, it is those individuals who developed “noise-tolerant” behavior that performed better. The lesson learned from this result is that interventions aimed to enhance complex skill acquisition should focus on
guiding learners towards “noise-tolerant” solutions. To further test this notion, we developed an intervention for guiding learners to the desirable solution in a rhythmic ball bouncing task. Due to its rhythmic nature, this task affords dynamically stable solutions that are robust to small errors and noise, a strategy that is independent from actively correcting error. Based on the task model implemented in a virtual environment, a subtle but systematic time-shift was implemented to change the range of ball–racket contacts that achieved dynamically stable solutions. Consistent with Manley et al. (2014) who demonstrated that interventions that facilitate simple skill learning by adding noise may not be appropriate for guiding subjects toward robust solutions, this study also showed that adding noise was not able to guide subjects towards the dynamically stable solutions.

While prior attempts to enhance learning of this particular complex rhythmic skill have fallen short (Marchal-Crespo et al., 2014; Morice et al., 2007), the novel guidance approach developed in this study was able to expedite learning. Although the subjects who practiced with this modification did not achieve better task performance than under normal conditions, the strategy they learned was more desirable in terms of control: they adopted solutions that were more dynamically stable and hence more immune to potential perturbations.

This insight is yet another important consideration when guiding acquisition of complex skills. By definition, tasks with redundancy have multiple successful solutions, but not all solutions are equal from a control perspective. Thus, measuring task success alone is not enough to discern the effectiveness of an intervention.

### 8.2.2 Importance of Persistence After Guidance

Irrespective of the task redundancy, an important consideration when developing interventions is that the learned behavior persists when the intervention is removed.
Therefore, the two studies on guidance also tested the retention of the learned behavior after guidance was terminated. For rhythmic ball bouncing, we showed that the learned behavior continued upon the removal of the guidance, even though only few sessions were tested. With the throwing task, persistence was tested even further. In the non-redundant task version, the reduced neuromotor noise was maintained for another five days after the manipulation was removed. This level of scrutiny into persistence of learned behavior is critical for extending guidance interventions into rehabilitation practice. However, it is often omitted in the experimental protocols on motor learning, probably due to practical considerations. More tests of retention of acquired skills also after periods without practice are needed (M. Abe et al., 2011; Park et al., 2013; Park & Sternad, 2015).

### 8.3 Psychological Influences on Motor Skill Learning

As the complexity of a motor task increases, the influence of cognitive, emotive, and other psychological processes on motor performance also increases. Hence, the effects of attention, motivation, and perceived effort need to be considered when assessing complex skill learning. As experimental rigor and replication is mandatory, it is necessary to know how such factors could potentially affect experimental results. One seemingly remote example is the gender of the experimenter, which has been shown to have differential effects on subject behavior under certain conditions (Rumenik, Capasso, & Hendrick, 1977).

In rehabilitation, equipping clinicians with an understanding of how psychological factors affect learning could allow them to effectively use such cues to enhance skill learning (Hall et al., 2010; Maclean & Pound, 2000). In two studies, we showed that a
verbal instruction intended to invoke a stereotype for females increased females’ motivation to perform better, according to the motivational mere effort account. It is not necessarily surprising that this instruction influenced performance given that motivation or engagement of the learner is integral for learning (M. F. Levin et al., 2015; Nudo, Milliken, Jenkins, & Merzenich, 1996). Nevertheless, the novel result from this study is that the same instruction could both assist and suppress skilled performance at different stages of learning. Moreover, the same instruction, delivered just before practice, decreased rhythmic ball bouncing performance, but enhanced performance in the discrete version of the same task. This research contributed to our understanding of that differential effect of instruction was likely due to the fact that the underlying mechanisms of learning and control are different for rhythmic and discrete tasks as demonstrated by behavioral, modeling, and neuroimaging results (Hogan & Sternad, 2007; Howard et al., 2011; Ikegami et al., 2010; Ronsse et al., 2009; Schaal et al., 2004; Sternad, Dean, & Schaal, 2000; Sternad et al., 2013).

A more explicit understanding how to avoid negative motivation and even exploit these psychological cues to enhance and not inadvertently debilitate learning would be beneficial for clinicians. Moreover, the lack of response to such cues might also signal the cognitive or emotional impairments in patients with neurological disorders or injuries. A particularly complex population in this respect is individuals with autism spectrum disorder. While results in this thesis highlight that interpersonal issues are important for rehabilitation, much more research is still needed. The research in this thesis also demonstrated that such cognitive effects cannot always be observed in course-grained performance measures but nonetheless influence behavior. Therefore future research on
the influence of verbal and cognitive cues should embrace motor skill paradigms that allow for more direct measures of motor control and behavior.

8.4 Future Directions for Improving Neurorehabilitation

While this thesis contributed behavior-based insights to our understanding of complex skill learning, motor rehabilitation also needs complementary research in the fields of neuroscience and neurophysiology. The benefit of behavioral research is that the insights can be directly translated into rehabilitation practice, but, evidently, these insights should be supported by neurophysiological mechanisms. Specifically, understanding the neural mechanisms of skill learning would allow us to identify neural deficits from overt behavior and inform how to train patients to either compensate or regain lost control through behavioral intervention.

8.4.1 Identifying Neural Mechanisms of Motor Control

Therefore, capitalizing on richer behavioral measures of control and learning, the next step is to reveal the neurophysiological underpinnings of such control and learning mechanisms. For example, one study from this thesis found that noise in motor behavior can be reduced by increasing the challenge for reward. While computational modeling results presented a rationale for this reduction of noise, it remained elusive how neuromotor noise reduced in the central nervous system. While we speculated about the co-activation of muscles and neuromodulation as potential mechanisms, more direct research is needed to bridge the gap between the neural and behavioral mechanisms of this observation.

Results from this line of research would also have immediate practical applications in rehabilitation. Currently, there are common neurological disorders with unknown
etiolologies. This lack of understanding between neurophysiological impairments and behavioral symptoms makes diagnosing, and consequently treating such disorders difficult. Are motor symptoms the result of neural impairments in perception, motor action, or in the integration process from perception to motor action? Complementing the behavioral research as presented in this thesis with neuroimaging and neurophysiological techniques could address this issue. However, it is important to recognize that identifying the neural underpinnings of motor symptoms is rife with challenges. For instance, neuroimaging studies have shown that neural dysfunction in children with developmental coordination disorder is not localized in the brain, but rather distributed over multiple areas including the cerebellum as well as parts of the frontal, temporal, and parietal lobules (Kashiwagi, Iwaki, Narumi, Tamai, & Suzuki, 2009; Peters, Maathuis, & Hadders-Algra, 2013; Zwicker, Missiuna, Harris, & Boyd, 2011, 2012). Thus, pinpointing the exact neural origins of specific motor symptoms is impossible in many cases. Although, the use of behavioral measures like those presented in this thesis combined with more advanced neural measures should make this feat less problematic.

8.4.2 Further Understanding Neuroplasticity During Motor Skill Learning

Additionally, a more detailed understanding of neuroplastic processes, during both healthy learning and motor recovery is needed. For instance, what neural substrates play a role in learning, and importantly, how do these roles change over the course of learning? Picard and colleagues recently found that years of practice resulted in more metabolically efficient generation of neuronal activity in the motor cortex during skilled performance in monkeys (Picard et al., 2013). Kawai and colleagues subsequently demonstrated in rats that while the motor cortex is critical for learning new skills, it is not required for executing learned skills (Kawai et al., 2015). While these animal studies shed
light onto the role of the motor cortex, the contributions of motor cortex and other neural substrates during learning and executing skills are still not fully understood, especially in humans. For instance, contrary to the results of (Kawai et al., 2015), human studies suggest that the motor cortex is needed for the retention of motor memories, whereas the cerebellum is required for learning (Galea et al., 2015; Galea, Vazquez, Pasricha, de Xivry, & Celnik, 2011). Just as the discovery of neuroplasticity provided powerful motivation for new directions in neurorehabilitation, a more fine-grained understanding of neuroplasticity will lead to more effective rehabilitation practices in the future.

8.4.3 Identifying Impairments in Learning Versus Impairments in Control

Another major open issue for neurorehabilitation research to address is whether the overt motor symptoms are a result of deficits in learning or due to loss of control, or even a combination of both. After stroke, for example, it is currently unclear whether mechanisms of neuroplasticity are impaired. On the one hand, Krakauer and colleagues state that there is insufficient evidence to suggest that the ability to learn or learn skills is impaired after stroke (Kitago & Krakauer, 2013; Krakauer, 2015). On the other hand, (M. F. Levin, 2011) suggests that cognitive deficits resulting from an acquired brain injury may indeed affect learning mechanisms. As demonstrated in this thesis, even subtle cognitive cues can influence skill acquisition. Realistically, the answer to this question most likely depends on many factors such as the size, type, and location of the affected neural substrate and therefore varies across individuals. This again emphasizes the critical need to gain more precise understanding of the neuromechanisms of learning and control.

At the same time, however, recent behavioral research aimed at addressing this question demonstrates the power of the fine-grained measures discussed in this thesis. For instance, one way to determine if motor impairments result from deficits in control or
learning is to examine long-term retention (Krakauer, 2015). Examining learning with the same virtual throwing task used in this thesis, Pendt and colleagues showed that patients with Parkinson’s disease did not have impairments in retention (Pendt, Maurer, & Müller, 2012; Pendt, Reuter, & Müller, 2011). However, fine-grained measures, similar as in this thesis demonstrated that initially large performance decrements on each practice day were likely the result of control deficits. In contrast, these patients did not have impairments in the fine-tuning processes of learning. Another study by Chu and colleagues using the virtual throwing task demonstrated that children with dystonia were not necessarily impaired in complex skill learning, despite their high variability (Chu, Park, Sanger, & Sternad, 2016). While their overall task performance was inferior to healthy subjects, dystonic children developed timing-insensitive arm trajectories to compensate for their variability and thereby improved task performance similar to healthy subjects.

Further research combining both fine-grained measures of behavior and neurophysiology is likely the key for addressing the open questions regarding neuromotor processes and recovery. Results from such research undoubtedly will have a major impact on creating the most effective form of rehabilitation for a particular patient population.

8.5 Future of Virtual Reality and Robotic Rehabilitation

Virtual and robotic technologies have been paramount in aiding our understanding of motor control over the past three decades. In the past, these technologies have been used to study the learning and control of relatively simple movements and tasks (Schmidt & Lee, 2011; Shadmehr & Mussa-Ivaldi, 1994; Wolpert et al., 2011). Sensibly, these
same motor behaviors were also used when robotic and virtual technologies were introduced into motor rehabilitation. Now that robotic-assisted therapy has demonstrated its potential in several large-scale clinical trials, efforts should be focused on how to increase its effectiveness for improving functional outcome. Thus, this thesis demonstrated how virtual environments can be used to enhance our understanding of more complex skill learning and simultaneously help design new research approaches and interventions to accelerate behavioral changes in patient populations.

Currently, the manipulation of feedback in virtual environments is one approach used in virtual rehabilitation (Holden, 2005; Sveistrup, 2004). However virtual environments afford many more opportunities to guide behavior that are not realizable with real-world tasks, particularly by manipulating task physics in a theoretically-motivated manner. This thesis research demonstrated how subjects still exploit the dynamically stable solutions, despite a manipulation of the task physics. Using a similar ball bouncing paradigm, Morice et al. (2007) also showed that subjects learn new attractor solutions created by a phase delay. The results of these studies not only provided further evidence that humans seek dynamically stable solutions, they also point towards potential interventions to guide learning in rehabilitation.

Despite their advantages, some considerations remain to be addressed. Currently, it is assumed that the strategies learned in the virtual environment are the same strategies humans engage in the real world. This equation is by no means guaranteed, or even tested and supported. Thus, future work needs to test whether the process of learning a virtual task is the same as learning a real world task. To justify the use virtual and robotic technologies in rehabilitation, research is needed to demonstrate that training on a virtual
task transfers into the performance of similar real world tasks (Bossard, Kermarrec, Buche, & Tisseau, 2008; Hyltander, Liljegren, Rhodin, & Lööroth, 2002; Rose et al., 2000). Lastly, prospective studies should demonstrate that training patients on virtual experimental tasks leads to functional improvements in performing activities of daily living. Such research would solidify the value of clinical adoption of virtual and robotic rehabilitation and have great influence on patients’ lives.

8.6 Concluding Remarks

Virtual reality and robotic devices not only offer great potential for rehabilitation, they also created a versatile platform for measurement and experimental manipulations. The research in this thesis harnessed virtual technology and contributed to our understanding complex skill learning, while simultaneously identifying new avenues for motor rehabilitation. However, there is still a great deal of behavioral and neurophysiological research needed to enhance the efficacy of future motor rehabilitation practices.

Moving forward, this thesis emphasized that more attention should be paid to complex skills, not only because they are relevant to everyday life, but also because tasks with high degrees of redundancy afford different learning strategies than tasks without redundancy. This is important given that rehabilitation should eventually transition towards enhancing complex skills as patients become more capable over the course of therapy. This is especially true for less impaired patient populations, that nevertheless have trouble executing activities of daily living, such as those with developmental coordination disorder or mild stroke.
A next step for research on skills is to gain a better understanding how humans learn to interact with objects that have complex dynamics (Hasson, Shen, et al., 2012; Hasson & Sternad, 2014; Nasseroleslami, Hasson, & Sternad, 2014). Additionally, the influence of cognitive processes should be considered and quantified during skill learning, especially as task complexity increases (Lee et al., 1994). Again, virtual tasks provide a perfect platform for this type of research as they afford ecologically relevant and stimulating environments, while maintaining systematic control (Sveistrup, 2004). Furthermore, mathematical models and analyses of virtual tasks afford more fine-grained measures of behavior and control, as task success alone may not be enough to determine the effectiveness of an intervention.

Lastly, behavioral researchers and neurophysiologists alike need to begin linking behavioral learning mechanisms to neural processes, especially now that a number of non-invasive brain measurement tools have become available. Pursuing these future avenues of research will guarantee better lives for patients, which is the main goal of motor rehabilitation.
References


Chalabaev, A., Brisswalter, J., Radel, R., Coombes, S. A., Easthope, C., & Clément-


Clark, R. A., Pua, Y.-H., Oliveira, C. C., Bower, K. J., Thilarajah, S., McGaw, R., …


Donchin, O., Francis, J. T., & Shadmehr, R. (2003). Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: Theory


Golomb, M. R., McDonald, B. C., Warden, S. J., Yonkman, J., Saykin, A. J., Shirley, B.,


Huber, M. E., Rabin, B., Docan, C., Burdea, G. C., AbdelBaky, M., & Golomb, M. R.


Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. a, Ruparel, K., …


Jørgensen, H. S., Nakayama, H., Raaschou, H. O., Vive-Larsen, J., Støier, M., & Olsen,


dysfunction in developmental coordination disorder: a functional MRI study. 


http://doi.org/10.1093/brain/awg079


http://doi.org/10.2106/JBJS.G.01483


http://doi.org/10.1371/journal.pone.0086580


http://doi.org/10.1016/j.cub.2015.05.038


Rose, F. D., Attree, E. A., Brooks, B. M., Parslow, D. M., Penn, P. R., & Ambhaipahan,


Shadmehr, R., & Moussavi, Z. M. (2000). Spatial generalization from learning dynamics


Towfighi, A., & Saver, J. L. (2011). Stroke declines from third to fourth leading cause of


Appendix A. Model for Rhythmic Ball Bouncing

The vertical position of the virtual ball $x_b$ between the $k$th and the $k+1$th racket-ball impact follows ballistic flight:

$$x_b(t) = x_b(t_k) + v_b^+(t - t_k) - g/2 (t - t_k)^2$$

where $t_k$ is the time of the $k$th ball-racket impact, $v_b^+$ is the velocity of the ball just after impact, and $g$ is the acceleration due to gravity (9.81 m/s$^2$). To determine the ball velocity just after impact $v_b^+$, an instantaneous impact is assumed as follows:

$$\alpha (v_b^-(t_k) - v_r^-(t_k)) = -(v_b^+(t_k) - v_r^+(t_k))$$

where $v_b$ and $v_r$ are the racket and ball velocities just before (-) and after (+) impact, and the energy loss at the collision is governed by the coefficient of restitution. The mass of the racket is assumed to be much larger than the mass of the ball, such that the racket velocity does not change during impact:

$$v_r^-(t_k) = v_r^+(t_k) = v_r(t_k).$$

Thus the ball velocity just after impact was determined by:

$$v_b^+(t_k) = (1 + \alpha) v_r(t_k) - \alpha v_b^-(t_k).$$

The racket and ball system can be modeled as a continuous dynamical system with sinusoidal racket motion. With this assumption, a discrete map can be derived based on two state variables, the ball velocity just after impact $v_b^+$ and the racket phase at impact $\theta_k$. Local linear stability analysis of this discrete map identifies a period-1 attractor, when racket acceleration at impact $a_r$ satisfies the inequality (Schaal et al., 1996; Sternad & Dijkstra, 2004):

$$-2g \frac{(1 + \alpha^2)}{(1 + \alpha)^2} < a_r < 0$$
Simulations of the ball bouncing map illustrate that when the impact occurs during negative racket acceleration of the upward racket swing, the ball exhibits stable period-1 behavior (Figure 5.2B). The map possesses other attractors besides the period-1 attractor, including a “sticking” behavior, where the ball follows the racket trajectory. This map exhibits “sticking” behavior when the ball-racket impact occurs during positive racket acceleration (Figure 5.2C).
Appendix B. Model for Rhythmic Ball Bouncing with Time-Dependent Manipulation

Under normal conditions, as in Experiment 1, the ball velocity immediately after impact was determined by:

\[ v_b^+(t_k) = (1 + \alpha) v_r(t_k) - \alpha v_b^-(t_k). \]

In Experiments 2 and 3, the racket velocity at impact \( v_r \) was set equal to racket velocity 50 ms before the time of impact \( t_k \). Thus the ball velocity just after impact was determined by:

\[ v_b^+(t_k) = (1 + \alpha) v_r(t_k - .05) - \alpha v_b^-(t_k). \]

As in the unperturbed map, the ball exhibits “sticking” behavior if the impact occurs during the positive racket accelerations (Figure 5.8A). During negative racket acceleration, however, initial impact phases that previously led to stable period-1 behavior in the unperturbed map now produce “sticking” behavior; only the more negative racket acceleration impact phases continue to produce stable period-1 behavior (Figure 5.8B). In fact, the time-dependent manipulation causes the domain of attraction for period-1 to shift by \( .05 \omega \) radians on the sinusoidal racket trajectory, where \( \omega \) is the angular frequency (Figure 5.8C).
Appendix C. Model for Discrete Ball Bouncing

The advantage of simulating the ball movement in a virtual environment is that it provides the explicit mathematical equations for motor execution and task performance. In contrast to using a real ball, there are no uncontrolled confounds, such as wind, friction, or spin of the ball.

After impact, the vertical position of the ball is determined by Eq. 1, where $t_{\text{impact}}$ is the time of ball-racket impact, $r_p$ is racket position at impact, and $b_v^+$ is the ball velocity just after impact.

$$b_p(t) = r_p + b_v^+(t - t_{\text{impact}}) - \frac{g(t - t_{\text{impact}})^2}{2}$$  \hspace{1cm} (1)

The ball is modeled as a point mass that impacts a racket, modeled as a planar surface. At the start of the trial, the ball rolls horizontally along the target line. Once the ball reaches the center of the screen, it drops vertically towards the racket from a starting height ($b_{p,0}$). Before the ball impacts the racket, its vertical position ($b_p$) is determined by Eq. 1, where $t$ is the time since the ball started to drop and $g$ is the acceleration due to gravity (9.81 m/s$^2$).

With the assumption of instantaneous impact, the ball velocity just after impact ($b_v^+$) is determined by Eq. 2, where $b_v$ and $r_v$ are the ball and racket velocities just before impact, and the energy loss at the collision is governed by the coefficient of restitution $\alpha$.

The ball velocity just before impact ($b_v$) is determined by Eq. 3.
In this experimental task, participants are instructed to bounce the ball such that the maximum height of the ball is equal to the target line. The maximum height of ball after impact is determined by Eq. 4. The task performance measure, error, is defined as the signed difference between the target height and the maximum ball height after impact (Eq. 5).

\[
max(b_p) = \frac{-b_v^+^2}{-2g} + r_p \quad \text{(4)}
\]

\[
Error = max(b_p) - \text{Target} \quad \text{(5)}
\]

Substituting Eqs. 2-4 into Eq. 5 yields an expression for error, which depends on two variables measured from the participant’s motor execution: racket position \(r_p\) and racket velocity \(r_v\) at impact (Eq. 6). The remaining variables are constants, which in this experiment had the following values: \(\alpha = 0.8\), \(b_{p,0} = 1.0275\)m, and \(\text{Target} = 1\)m.

\[
Error = \frac{-((1 + \alpha)r_v - \alpha b_v)^2}{-2g} + r_p - \text{Target} \quad \text{(6)}
\]
Appendix D. Statistical Comparisons of Baseline Participants with No-Threat (NT) Participants

To ensure that the NT instruction did not influence performance compared to the baseline condition, a 2 (NT vs Baseline) x 2 (Male vs Female) x 12 (Blocks) ANOVA was conducted on each dependent measure. The results summarized in Table D.1 and Figure D.1 below indicated that there was no influence of the NT instruction on any of the dependent measures.

<table>
<thead>
<tr>
<th>Table D.1. Results of statistical analyses between baseline vs no threat (NT) participants.</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of Successful Bounces</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Block</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Condition</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Block x Gender</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Block x Condition</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Gender x Condition</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Block x Gender x Condition</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The shaded boxes indicate a significant effect or interaction.
Figure D.1. Comparison of baseline vs no-threat (NT) participants. (A) Task success, (B) median error, and (C) IQR error over blocks. Each point represents the group average of the dependent measures per block, and the error bar represents the standard error across participants in each group.
Appendix E. Statistical Analyses of Baseline Participants with Stereotype Threat (ST) Participants

To test if ST instruction enhanced the performance compared to the baseline group, a 2 (ST vs Baseline) x 2 (Male vs Female) x 12 (Blocks) ANOVA was conducted on each dependent measure. The results summarized in Table E.1 and Figure E.1 below indicated that the ST instruction did enhance performance as measured by the percentage of successful bounces and IQR of error. These results are similar to those of the Stereotype Experiment reported in the main text.

Table E.1. Results of statistical analyses between baseline vs stereotype threat (ST) participants.

<table>
<thead>
<tr>
<th></th>
<th>% of Successful Bounces</th>
<th>Median Error</th>
<th>IQR of Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>$F(11, 495) = 19.34$</td>
<td>$F(11, 495) = 3.35$</td>
<td>$F(11, 495) = 46.65$</td>
</tr>
<tr>
<td></td>
<td>$p&lt;.001$</td>
<td>$p = .022$</td>
<td>$p &lt; .001$</td>
</tr>
<tr>
<td>Gender</td>
<td>$F(1,45) = .72$</td>
<td>$F(1,45) = .006$</td>
<td>$F(1,45) = 1.70$</td>
</tr>
<tr>
<td></td>
<td>$p = .40$</td>
<td>$p = .94$</td>
<td>$p &lt; .20$</td>
</tr>
<tr>
<td>Condition</td>
<td>$F(1,45) = 79.26$</td>
<td>$F(1,45) = .38$</td>
<td>$F(1,45) = 14.05$</td>
</tr>
<tr>
<td></td>
<td>$p &lt; .001$</td>
<td>$p = .54$</td>
<td>$p &lt; .001$</td>
</tr>
<tr>
<td>Block x Gender</td>
<td>$F(11, 495) = 1.73$</td>
<td>$F(11, 495) = .45$</td>
<td>$F(11, 495) = 2.41$</td>
</tr>
<tr>
<td></td>
<td>$p = .095$</td>
<td>$p = .71$</td>
<td>$p &lt; .067$</td>
</tr>
<tr>
<td>Block x Condition</td>
<td>$F(11, 495) = 2.27$</td>
<td>$F(11, 495) = .27$</td>
<td>$F(11, 495) = .85$</td>
</tr>
<tr>
<td></td>
<td>$p = .025$</td>
<td>$p = .84$</td>
<td>$p = .47$</td>
</tr>
<tr>
<td>Gender x Condition</td>
<td>$F(1,45) = .33$</td>
<td>$F(1,45) = 1.27$</td>
<td>$F(1,45) = .12$</td>
</tr>
<tr>
<td></td>
<td>$p = .57$</td>
<td>$p = .27$</td>
<td>$p = .73$</td>
</tr>
<tr>
<td>Block x Gender x Condition</td>
<td>$F(11, 495) = 2.027$</td>
<td>$F(11, 495) = .45$</td>
<td>$F(11, 495) = .96$</td>
</tr>
<tr>
<td></td>
<td>$p = .046$</td>
<td>$p = .71$</td>
<td>$p = .41$</td>
</tr>
</tbody>
</table>

The shaded boxes indicate a significant effect or interaction.
Figure E.1. Comparison of baseline vs stereotype threat (ST) participants. (A) Task success, (B) median error, and (C) IQR error over blocks. Each point represents the group average of the dependent measures per block, and the error bar represents the standard error across participants in each group.
Appendix F: Validity and Reliability of Kinect Skeleton for Measuring Shoulder Joint Angles

A common solution in the virtual rehabilitation community has been to use low-cost sensors initially intended for commercially available video games. The low cost of these sensors is especially appealing; however, these gaming sensors and their complementary motion analysis software were not originally designed to capture movements with accuracy necessary in research and clinical applications for patient populations with impaired motor abilities. Thus, precautions must be made to ensure that the sensors used in VR rehabilitation systems can accurately measure motor behavior, especially before any deployment for in-home rehabilitation systems. Not only is high measurement accuracy required to track and assess patient progress, it is also vital to ensure that patients perform the correct movements without compensatory movements or exceeding prescribed range of motion. Therefore, this study examined the validity and reliability of one of the more popular low cost VR sensors, the Microsoft Kinect. This study was intended as the first step in the development of an in-home rehabilitation program for patients after shoulder operation (Huber, Leeser, & Sternad, 2013). However, the inadequate accuracy of the Kinect system terminated this project plan. A Provost Tier I grant from Northeastern University and a grant from The Mathworks funded this research, which involved Dagmar Sternad, Miriam Leeser and Amee Seitz.

F.1 Abstract

Objective: To test the reliability and validity of shoulder joint angle measurements from the Microsoft Kinect™ for virtual rehabilitation. Design: Test-retest reliability and concurrent validity study. Setting: Motion analysis laboratory. Participants: A convenience sample of 10 healthy adults. Methods: Shoulder joint angle was as concurrently assessed with two trials in 4 static poses using: (1) the Kinect, (2) 3D motion analysis system, and (3) clinical goniometer. All poses were captured with the Kinect from the frontal view. The two poses of shoulder flexion were also captured with the Kinect from the sagittal view. Main Outcome Measures: Absolute and relative test-retest reliability of the Kinect for measuring shoulder angles was determined in each pose with intraclass correlation coefficients (ICCs), standard error of the measure (SEM), and minimal detectable change (MDC). The 95% limits of agreement (LOA) between the Kinect and the standard methods for measuring shoulder angle were computed to determine concurrent validity. Results: While the Kinect provided to be highly reliable (ICC = 0.76-0.98) for measuring shoulder angle from the frontal view, the 95% LOA between the Kinect and two measurement standards were greater than ± 5° in all poses for both views. Conclusions: Before the Kinect is used to measure movements for virtual rehabilitation applications, understanding its limitations in precision and accuracy of measuring joint motion in all planes is imperative.

F.2 Introduction

The use of virtual reality technology for rehabilitation, or virtual rehabilitation (VR), provides several advantages over conventional therapy, including an increased
capacity to quantitatively measure motor performance, deliver real-time performance feedback, and enhance patient motivation. By exploiting the latest commercial game technologies (Deutsch, Borbely, Filler, Huhn, & Guarrera-Bowlby, 2008; Golomb et al., 2010; Huber et al., 2010; Morrow et al., 2006), VR systems are being developed at increasingly low costs, making them especially ideal for in-home therapy.

Much of the current research using in-home VR is aimed towards aiding patients with neurological disorders (Deutsch et al., 2008; Golomb et al., 2010; Huber et al., 2010; Morrow et al., 2006; Pompeu et al., 2014). However, these systems also have the potential to improve physical therapy for patients with musculoskeletal disorders. With an in-home VR system, a clinician can ensure that the patient is performing exercises correctly and reaching functional goals targeted in specific post-operative timeframes that allow proper joint healing. Yet, despite these advantages, the use of VR for post-operative joint therapy is currently very limited.

The Microsoft Kinect™, one of the more popular gaming sensors, is an ideal sensor for a VR system designed for post-operative joint rehabilitation. Like all gaming technology, however, the Kinect was not developed with intention of clinical use. Thus, the accuracy of Kinect measurements must be thoroughly evaluated for movements of interest before clinical application. It is disconcerting for a study using the Kinect for VR to claim that the sensor’s validity and reliability has been previously established, yet cite studies that have assessed the validity and reliability of different Kinect measurements (Pompeu et al., 2014). For instance, it is misleading to cite the accuracy of the depth image to assure accuracy of skeletal data from the Microsoft Kinect for Windows™ Software Development Kit (SDK) (Shotton et al., 2011).
With the ultimate goal of developing a VR system for post-operative shoulder therapy, this study aimed to assess the reliability and validity of skeletal data from the Kinect for Windows SDK for measuring precise shoulder angles. Previous work has found that the skeletal data from the Kinect SDK can be used to accurately measure shoulder range of motion (ROM) (Bonnechère et al., 2014), but its accuracy for measuring exact shoulder angles has not been investigated. A prior study by Fernandez-Baena, Susin, and Lligadas (2012), which also examined accuracy of the Kinect for measuring shoulder ROM, observed average errors between 8 and 14 degrees in shoulder angle trajectories. C.-Y. Chang et al. (2012) similarly observed large errors in the tracking the shoulder position with the Kinect. However, no formal analyses of validity and reliability were conducted in these studies. Furthermore, these studies used the skeletal data from OpenNI (Primesense), which is different from the skeletal data from the Kinect SDK. Nonetheless, the findings of C.-Y. Chang et al. (2012) and Fernandez-Baena et al. (2012) identify the need to assess accuracy for measuring exact shoulder angles with the Kinect skeletal data.

The present study measured a number of different shoulder angles while participants held a series of static poses. These poses consisted of shoulder configurations commonly used in post-operative shoulder rehabilitation, including one pose where the shoulder was occluded from the view of the Kinect. The shoulder angle measures from three data acquisition systems, the Kinect, 3D motion analysis system (gold standard), and goniometer (clinical standard), were compared.
F.3 Methods

F.3.1 Participants

A convenience sample of 10 asymptomatic adults with no known shoulder pathology (6 females and 4 males, mean age 22.1 ± 0.9 years) from Northeastern University took part in the experiment. All participants gave informed written consent before the experiment. The experimental protocol was approved by the University’s Institutional Review Board.

F.3.2 Procedure

With the Kinect in the frontal view, each participant held the following static poses, two repetitions each, in a randomized order: Flexion to 90°, Flexion to Max, Abduction to 90°, and External Rotation to Max at 0° Abduction (Figure F.1). Two additional repetitions of both the Flexion to 90° and Flexion to Max poses were measured with the Kinect in the sagittal view. The sagittal view poses were added as pilot work revealed that during Flexion to 90°, the shoulder joint was occluded from the Kinect from the frontal view. Thus, the two poses of shoulder flexion were repeated from the sagittal view to determine if shoulder flexion measurements were reliable and valid from this vantage point. Half of the participants performed the shoulder motion with the dominant arm and the other half with the non-dominant arm, also randomized.

The 90° poses (Flexion to 90° and Abduction to 90°) were set using the goniometer. For the Max poses (Flexion to Max and External Rotation to Max at 0° Abduction), participants were instructed to rotate to their maximum capability. Once the pose was set, measurements were simultaneously recorded using three modalities: (1) Kinect for Windows (Microsoft), (2) 3D motion analysis system (Acension Trakstar) and (3) a
blinded goniometer. The pose was reset for each repetition. It is feasible that there were differences between repetitions in the max values. However, this procedure was consistent with current practice. We reduced the likelihood of variations in ROM between trials of a pose by utilizing healthy, pain-free subjects.

F.3.3 Data Capture and Processing

F.3.3.1 Kinect

The skeleton data captured from the Kinect for Windows SDK for each pose consisted of the 3D positions of 20 joints. The positions of shoulder and elbow joints relative to the trunk were used to measure the angles of shoulder flexion and abduction (in degrees), while the positions of the elbow and hand relative to the trunk were used to measure the angle of external rotation. Skeletal data from the Kinect for Windows SDK was accessed and analyzed in MATLAB (The Mathworks).

F.3.3.2 Goniometer

A standard 12-inch goniometer was modified so that the measures were blinded to the examiner. Goniometric measurements of shoulder joint motions were performed using standardized methods (Jain, Wilcox, Katz, & Higgins, 2013). Once the goniometer was aligned to the shoulder motion by the examiner, a second examiner read and recorded the measurement in degrees. Goniometric measures of the shoulder have demonstrated excellent reliability (Hayes, Walton, Szomor, & Murrell, 2001; Riddle, Rothstein, & Lamb, 1987).

F.3.3.3 3D Motion Analysis

The Ascension Trakstar electromagnetic-based motion analysis system (Ascension Technology Corporation) with a sampling rate of 240 Hz was used with Motion Monitor software (Innovative Sports Training, Inc.) to collect 3D kinematic data of the humerus.
and trunk. Electromagnetic receivers were secured with tape on the thorax over the spinous process of the T3 vertebrae and the posterior aspect of the distal humerus of the arm. Local coordinate axes systems for each segment were created using digitized anatomical landmarks on each segment as previously described following International Society of Biomechanics (ISB) recommendations (G. Wu et al., 2005). Euler angle sequences for humeral rotations were used to describe motion of the humerus relative to the thorax. Shoulder movements into the directions of abduction and flexion were described using elevation angles and external rotation with long axis rotation. All motions in flexion, abduction and external rotation were defined as positive for direct comparison to clinical goniometry measures. Previously reported RMS accuracy of this system was <1° (Ludewig et al., 2009).

F.3.4 Statistical Analysis

The intra-class correlation coefficient model 3,2 or ICC(3,2) was employed to determine the relative test-retest reliability of the Kinect for measuring shoulder angles (Shrout & Fleiss, 1979). Six ICC(3,2) values, representing the agreement of two trials for each pose from the respective views, were computed. The ICC(3,2) values were defined as: poor when below 0.20; fair from 0.21 to 0.40; moderate from 0.41 to 0.60; good from 0.61 to 0.80, and very good from 0.81 to 1.0 (Landis & Koch, 1977). The standard error of the measure (SEM) and the minimal detectable change (MDC) were calculated to establish the absolute reliability. SEM was defined as the standard deviation multiplied by the square root of the ICC subtracted from 1 (Stratford, Binkley, & Riddle, 1996), and MDC was calculated by multiplying the SEM by the square root of 2 (Stratford, Binkley, Solomon, et al., 1996).
The 95% limits of agreement (LOA) between the Kinect and the two measurement standards for shoulder, goniometer and 3D motion analysis, were computed for each pose to determine validity (Stratford, Binkley, Solomon, et al., 1996). To obtain the 95% LOA in each pose, first the mean of the two shoulder angle measurements from each method was calculated. Then the mean and standard deviation of differences between (1) Kinect and goniometer and (2) Kinect and 3D motion analysis measurements were computed. The 95% LOA were defined as the mean difference ±1.96 standard deviation of the differences, such that 95% of differences lay within these limits. If the 95% LOA were greater than ±5°, then the discrepancies between measurement systems were considered clinically significant (Bland & Altman, 1986). All statistical analyses were conducted with the IBM SPSS Statistics software (IBM Corporation).

F.4 Results

F.4.1 Test-Retest Reliability

Results of test-retest reliability of the Kinect for measuring shoulder angle with ICC, SEM, and MDC values are shown in Table F.1. From the frontal view, the Kinect had good to very good relative reliability for measuring shoulder angle in all four poses, as indicated by the high ICC(3,2) values. Small SEM and MDC values, indicating good absolute reliability, were observed for all poses from the frontal view except Flexion to 90°, where the shoulder joint was occluded from the Kinect camera. From the sagittal view, the Kinect had very good relative and absolute reliability in Flexion to 90°, but only fair relative and poor absolute reliability in Flexion to Max.

F.4.2 Concurrent Validity
The 95% LOA between the Kinect and goniometer and Kinect and 3D motion analysis system are shown in Table F.2. *Abduction to 90°* was the only pose in which the Kinect measures of shoulder angle were reasonably accurate when compared either to goniometer or 3D motion tracking system. Like the other poses, however, the 95% LOA for the discrepancy between systems exceeded ±5°, which we defined as clinically significant.

### F.5 Discussion

Using the skeletal data from the Kinect for Windows SDK, the Kinect tended to be highly reliability for measuring shoulder angle in most poses. Out of all poses, highest accuracy was achieved for the shoulder angle in the *Abduction to 90°* pose. This is consistent with the prior finding that the Kinect was accurate in measuring ROM during shoulder abduction (Bonnechère et al., 2014). However, the measurement discrepancies between the Kinect and the two measurement standards were clinically significant in all poses.

From the frontal view, Kinect struggled with measuring shoulder angle in the transverse plane (*External Rotation to Max at 0° Abduction*) and the sagittal plane (*Flexion to 90°* and *Flexion to Max*). (Bonnechère et al., 2014) reported similarly poor results using the Kinect to measure ROM during elbow flexion and knee flexion, which are both movements in the sagittal plane. Inaccurate measurements in the sagittal plane can be explained in part by error introduced from estimating the position of occluded joints, a problem typical of vision-based motion tracking systems. This was most evident from the poor absolute reliability and validity of the Kinect measurements in *Flexion to 90°* from the frontal view, where the shoulder joint was occluded by the arm.
In an attempt to circumvent this problem, *Flexion to 90°* and *Flexion to Max* were also measured from the sagittal view, but this also did not prove a viable solution. In *External Rotation to Max at 0° Abduction* and *Flexion to Max*, joint occlusion was not an issue, but the Kinect-measured shoulder angles were still inaccurate. This major limitation poses a problem for measuring simple shoulder movements as well as complex, full-body movements.

**F.6 Conclusions**

While the skeletal data of Kinect for Windows SDK may be accurate for commercial gaming purposes, this study revealed a significant concern for using these data to measure shoulder motion when precise shoulder angle measurements are required. In this study, only a limited number of shoulder configurations were examined. However, the large discrepancies in shoulder angle measurements should not be taken lightly. In fact, these results identify the need for further assessment of the Kinect’s accuracy for measuring a wider range of shoulder angles also in impaired movements.

VR for post-operative shoulder therapy is only one exemplary case where accurate shoulder angle measurements are needed. In fact, accurate shoulder angle measurements are required for most musculoskeletal and neurological rehabilitation protocols (C.-Y. Chang et al., 2012; Fernandez-Baena et al., 2012; Golomb et al., 2010; Lange et al., 2011). Before the Kinect can be used to measure movements in VR, understanding its limitations in precision and accuracy for measuring specific joint motions is imperative.
Figure F.1. Four static poses with Kinect skeleton overlayed in the respective views from the Kinect camera.
Table F.1. Test-retest reliability results of the Kinect for measuring shoulder angle.

<table>
<thead>
<tr>
<th>Kinect View</th>
<th>Pose</th>
<th>ICC</th>
<th>Mean</th>
<th>SEM</th>
<th>MDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front</td>
<td>Abduction to 90°</td>
<td>0.76</td>
<td>90.1°</td>
<td>2.5°</td>
<td>3.5°</td>
</tr>
<tr>
<td></td>
<td>External Rotation to Max at 0° Abduction</td>
<td>0.98</td>
<td>65.8°</td>
<td>3.7°</td>
<td>5.2°</td>
</tr>
<tr>
<td></td>
<td>Flexion to 90°</td>
<td>0.85</td>
<td>73.7°</td>
<td>12.2°</td>
<td>17.3°</td>
</tr>
<tr>
<td></td>
<td>Flexion to Max</td>
<td>0.95</td>
<td>162.2°</td>
<td>4.0°</td>
<td>5.6°</td>
</tr>
<tr>
<td>Sagittal</td>
<td>Flexion to 90°</td>
<td>0.84</td>
<td>86.7°</td>
<td>4.4°</td>
<td>6.2°</td>
</tr>
<tr>
<td></td>
<td>Flexion to Max</td>
<td>0.37</td>
<td>161.5°</td>
<td>24.2°</td>
<td>34.1°</td>
</tr>
</tbody>
</table>

*ICC, intra-class correlation coefficient; SEM, standard error of the measure; MDC, minimal detectable change.*

Table F.2. Concurrent validity results of the Kinect for measuring shoulder angle compared to goniometer and 3D magnetic tracking system.

<table>
<thead>
<tr>
<th>Kinect View</th>
<th>Pose</th>
<th>95% Limits of Agreement</th>
<th>Kinnect - Goniometer</th>
<th>Kinnect - 3D Magnetic Tracker</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LL</td>
<td>Mean</td>
<td>UL</td>
</tr>
<tr>
<td>Front</td>
<td>Abduction to 90°</td>
<td>-7.0°</td>
<td>-1.5°</td>
<td>4.1°</td>
</tr>
<tr>
<td></td>
<td>External Rotation to Max at 0° Abduction</td>
<td>-14.3°</td>
<td>1.3°</td>
<td>16.9°</td>
</tr>
<tr>
<td></td>
<td>Flexion to 90°</td>
<td>-52.1°</td>
<td>-16.6°</td>
<td>18.9°</td>
</tr>
<tr>
<td></td>
<td>Flexion to Max</td>
<td>-36.5°</td>
<td>-15.9°</td>
<td>4.7°</td>
</tr>
<tr>
<td>Sagittal</td>
<td>Flexion to 90°</td>
<td>-17.7°</td>
<td>-6.9°</td>
<td>3.8°</td>
</tr>
<tr>
<td></td>
<td>Flexion to Max</td>
<td>-44.5°</td>
<td>-15.0°</td>
<td>14.5°</td>
</tr>
</tbody>
</table>

*LL, lower limit = mean – 1.96 standard deviation; UL, upper limit = mean + 1.96 standard deviation*
Appendix G: Accuracy of Kinect for Measuring Shoulder Joint Angles in Multiple Planes of Motion

This study continued examination of the validity and reliability of one of the more popular low cost VR sensors, the Microsoft Kinect. A limitation of the previous study was that only a small number of shoulder configurations were examined. Therefore, this study expanded the number of examined shoulder configurations and again compared shoulder angle measurements from the Kinect with a 3D motion analysis system (gold standard) and goniometer measures (clinical standard).


G.1 Abstract

Virtual reality-based physical rehabilitation, or virtual rehabilitation, provides several advantages over conventional therapy. These include the capacity to provide patient-specific treatment that adapts with functional improvements over practice, obtain quantitative measures of progress, deliver real-time performance feedback through various modalities, and improve adherence by heightening patient motivation and entertainment. By exploiting commercially available gaming technology, virtual rehabilitation systems can even be developed at a low cost and conveniently used in the home for outpatient therapy. The Microsoft Kinect is one such gaming technology that has gained recent popularity within the virtual rehabilitation community. This highly advanced, and yet low cost sensor enables users to interact with the system by monitoring 3D body movements, making its clinical utility highly attractive to the rehabilitation
community. Before this technology can be translated into the clinical setting, it is paramount to ensure its precision and accuracy of measuring joint motion. The present study aimed to test the reliability and validity of upper extremity joint angle measurements with the Kinect for shoulder rehabilitation. Results indicate that while the Kinect is reliable for measuring shoulder joint angles, there are large discrepancies in measured shoulder angles from the Kinect compared to the gold (magnetic tracker) and clinical (goniometer) standards. Before the Kinect can be used to measure movements for virtual rehabilitation applications, understanding its limitations in precision and accuracy of measuring specific joint motions is imperative when measuring impaired movements.

G.2 Introduction

Virtual reality-based physical rehabilitation, or virtual rehabilitation, provides several advantages over conventional therapy, including an increased capacity to obtain quantitative measures to track motor performance, deliver real-time performance feedback through visual and auditory modalities, and enhance patient motivation and entertainment. By exploiting the latest commercial game technologies, such as the Nintendo Wii™ (Deutsch et al., 2008), the Microsoft Xbox™ (Morrow et al., 2006), and the Sony Playstation 3™ (Golomb et al., 2010; Huber et al., 2010), virtual rehabilitation systems are being developed at increasingly low costs, making them ideal for administering in-home therapy. These gaming technologies were not developed with intention of clinical use however. Thus, the accuracy of their measurements must be thoroughly evaluated before it can be used for rehabilitation. As with all new sensing systems, measurement accuracy for the specific measurement of interest should be thoroughly assessed before practical and, even more importantly, clinical use.
With the ultimate goal of developing a virtual rehabilitation system for post-operative shoulder therapy, this study aimed to assess the validity of the popular Microsoft Kinect™ to measure shoulder angles. Previous work of (Bonnechère et al., 2014) found the skeletal data from the Kinect for Windows SDK to be valid for measuring range of motion in shoulder abduction. However, this study did not investigate whether this data could also measure exact shoulder angles; furthermore, only one plane of shoulder movement was considered. Another study by Fernandez-Baena et al. (2012), which also examined accuracy of the Kinect for measuring shoulder range of motion, observed average errors between 8 and 14 degrees in shoulder angle trajectories. C.-Y. Chang et al. (2012) similarly observed large errors when tracking the shoulder position with the Kinect. However, no formal analyses of validity were conducted. Moreover, these two studies used the skeletal data from OpenNI (Primesense), which is different from the skeletal data from the Kinect for Windows SDK. Hence, these studies highlight the need to assess accuracy of measuring exact shoulder angles with the Kinect skeletal data.

Huber, Seitz, Leeser, and Sternad (2015) compared shoulder angle measurements from the Kinect for Windows SDK (v1.6) with 3D motion analysis system and goniometer measurement, which led to alarming results. The measurement discrepancies between the Kinect and the two measurement standards for shoulder angle were found to be clinically significant and exceeded error associated with traditional clinical measures of ±5°. The major limitation of this previous work was that only a small number of shoulder configurations were examined and the results identified the need for further assessment of the Kinect’s accuracy for a wider range of shoulder angles (Huber, Seitz, et
al., 2015). To address this need, this study expanded the number of examined shoulder configurations and again compared shoulder angle measurements from the Kinect with a 3D motion analysis system (gold standard) and goniometer measures (clinical standard).

G.3 Methods

G.3.1 Participants

Asymptomatic healthy participants were recruited using a sample of convenience from a university-based population. To be eligible, participants had to be at least 18 years of age and demonstrate normal shoulder and cervical spine active and passive range of motion without symptoms. Subjects were excluded if there was a known diagnosis of any systemic diseases that could affect the musculoskeletal system, known shoulder pathology, history of shoulder fracture or surgery, or implanted electronic or metal devices. Ten subjects (6 females and 4 males; mean age 21.4 ± 3.2 years) agreed to participate and signed the University's Institutional Review Board approved informed consent.

G.3.2 Experimental Protocol

Each participant held a series of 11 static poses shown in Figure G.1, each measured twice, but in fully randomized order. Three planes of motion were examined: external rotation in a 90 degree abducted position, flexion and abduction. Half of the participants performed the shoulder motion with the dominant arm and the other half with the non-dominant arm.

At the start of each trial, Experimenter 1 set the participant into the specific pose based on a jig created with angular gonimetric measurements. Once the pose was set, measurements were simultaneously recorded using three modalities: (1) Kinect, (2) 3D
motion analysis system (Acension Trakstar), and (3) a blinded goniometer as measured by Experimenter 2. The pose was reset for each new measurement.

G.3.4 Measurement Systems

The shoulder angle in each pose was simultaneously measured from three data acquisition systems: the Microsoft Kinect, goniometer (clinical standard), and 3D motion analysis system (gold standard).

G.3.4.1 Microsoft Kinect

The Kinect camera was placed approximately 2 m in front of participant, thus capturing the participant’s motion from the frontal view, which is the common setup. The skeleton data captured from the Kinect for Windows SDK for each pose consisted of the 3D positions of 20 joints. The positions of shoulder and elbow joints relative to the trunk were used to measure the angles of shoulder flexion and abduction (in degrees), while the positions of the elbow and hand relative to the humerus were used to measure the angle of external rotation. Skeletal data from the Kinect for Windows SDK was accessed and analyzed in MATLAB (The Mathworks).

G.3.4.2 Goniometer

A standard 12-inch goniometer was modified so that the measures were blinded to the experimenter. Goniometric measurements of shoulder joint motions were performed using standardized methods (Jain et al., 2013). Once the first experimenter aligned the goniometer to the shoulder motion, a second experimenter read and recorded the measurement in degrees. Goniometric measures of the shoulder have demonstrated excellent reliability (Hayes et al., 2001; Riddle et al., 1987).
G.3.4.3 3D Motion Analysis System

The Ascension Trakstar electromagnetic-based motion analysis system (Ascension Technology Corporation) with a sampling rate of 240 Hz was used with Motion Monitor software (Innovative Sports Training, Inc.) to collect 3D kinematic data of the humerus and trunk. Electromagnetic receivers were secured with tape on the thorax over the spinous process of the T3 vertebrae and the posterior aspect of the distal humerus of the arm. Local coordinate axes systems for each segment were created using digitized anatomical landmarks, as previously following the recommendations of the International Society of Biomechanics (ISB) (G. Wu et al., 2005). Euler angle sequences for humeral rotations were used to describe motion of the humerus relative to the thorax. Shoulder movements into the directions of abduction and flexion were described using elevation angles and external rotation with long axis rotation. All motions in flexion, abduction and external rotation were defined as positive for direct comparison to clinical goniometry measures. Previously reported RMS accuracy of this system was <1° (Ludewig et al., 2009).

G.3.5 Statistical Analyses

SEM was defined as the standard deviation multiplied by the square root of the ICC subtracted from one (Stratford, Binkley, & Riddle, 1996), and MDC was calculated by multiplying the SEM by the square root of two (Stratford, Binkley, Solomon, et al., 1996).

The 95% limits of agreement (LOA) between the Kinect and the two measurement standards for shoulder, goniometer, and 3D motion analysis were computed for each pose type (Flexion, Abduction, and External Rotation at 90° Abduction) to determine validity (Bland & Altman, 1986). To obtain the 95% LOA in each pose type, first the mean was
calculated from the two shoulder angle measurements from each method. Then the mean and standard deviation of differences between (1) Kinect and goniometer and (2) Kinect and 3D motion analysis measurements were computed. The 95% LOA were defined as the mean difference ±1.96 standard deviation of the differences, such that 95% of differences lay within these limits. If the 95% LOA were greater than ±5°, then the discrepancies between measurement systems were considered clinically significant.

G.4 Results

G.4.1 Test-Retest Reliability

Results of test-retest reliability of the Kinect for measuring shoulder angle with ICC, SEM, and MDC values are shown in Table G.1. Small SEM and MDC values, indicating good absolute reliability, were observed for all poses except for Flexion 90° and 0° External Rotation 90°. In Flexion 90°, the shoulder joint was occluded from the Kinect camera by the arm, and in 0° External Rotation 90°, the elbow joint was occluded from the Kinect camera. These two poses also had the lowest ICC values, indicating poor to fair relative reliability. The majority of the other poses had good to very good relative reliability, only two Abduction poses had only moderate relative reliability.

G.4.2 Validity

The 95% limits of agreement (LOA) between the Kinect and the two measurement standards for shoulder, goniometer, and 3D motion analysis for each pose type, Flexion, Abduction, External Rotation at 0°, and External Rotation at 90°, are given in Table G.2. For all pose types, the 95% LOA were greater than ±5°, indicating that the discrepancies between measurement systems were clinically significant. Figure G.2 shows the Bland-Altman plots for each pose type, which plots the difference between the two
measurements over the mean value of the two measurements (Bland & Altman, 1986). Figure G.2 shows that for all pose types the mean difference between measurements was non-zero, meaning that there was an offset between measurements systems in addition to the large variations in measurement differences.

G.5 Discussion

Using the skeletal data from the Kinect for Windows SDK, the Kinect tended to be highly reliable for measuring shoulder angle, except for two poses where either the shoulder or elbow joint was occluded from the Kinect camera. The prior study of Huber, Seitz, et al., (2015) attempted to circumvent this problem by measuring Flexion to 90° from the sagittal view, but this also was not a viable solution. As with most camera-based vision systems, occlusion remains a major problem. This limitation in the skeletal recognition algorithm not only presents a problem for measuring simple shoulder movements, but evidently even more so for quantifying complex, full-body movements.

While the Kinect had high test-retest reliability, it was not accurate for measuring exact shoulder angles as comparisons with goniometer and 3D magnetic tracking measurements demonstrated. The 95% LOA between the Kinect and the goniometer were the lowest for Abduction with a difference of ±11.7°. This is consistent with prior findings that revealed large measurement errors in shoulder angle trajectories during shoulder abduction (C.-Y. Chang et al., 2012; Fernandez-Baena et al., 2012). The largest 95% LOA between the Kinect and the goniometer were found in External Rotation at 90° Abduction. Bonnechère et al. (2014) reported similarly poor results using the Kinect to measure range of motion during elbow flexion and knee flexion, which are movements that occur in the same plane of motion as External Rotation at 90° Abduction.
G.6 Conclusions

While the skeletal data of Kinect for Windows SDK may be accurate for commercial gaming purposes, this study revealed a significant concern for using these data to measure joint angles when precise angle measurements are required. The large discrepancies in shoulder angle measurements should not be taken lightly. Virtual Rehabilitation for post-operative shoulder therapy is only one exemplary case where accurate shoulder angle measurements are needed. In fact, accurate shoulder angle measurements are required for most musculoskeletal and neurological rehabilitation protocols (Bonnechère et al., 2014; C.-Y. Chang et al., 2012; Golomb et al., 2010; Lange et al., 2011). Before the Kinect can be used to measure movements in virtual rehabilitation, it is imperative to understand its limitations in precision and accuracy for measuring specific joint motions.

Given the many other advantages of Kinect, efforts should be made to enhance Kinect’s precision for example by fusing the sensor data readings with those different sensors (Bó, Hayashibe, & Poignet, 2011; Destelle et al., 2014) or combine the data from multiple Kinect sensors (Asteriadis, Chatzitofis, Zarpalas, Alexiadis, & Daras, 2013; Berger, Ruhl, Schroeder, & Bruemmer, 2011). Thus despite its existing limitations, the Kinect remains a promising tool for virtual rehabilitation.
G.7 Figures and Tables

Figure G.1. Eleven poses with different shoulder configurations from one example subject. The skeletal data from the Microsoft Kinect for Windows SDK is overlaid in red.
Figure G.2. Bland-Altman plots for each pose type: (A) Flexion, (B) Abduction, and (C) External Rotation at 90° Abduction. The red circles denote results from comparing shoulder angle measures between the Kinect and goniometer. The blue triangles denote results from comparing shoulder angle measures between the Kinect and the 3D magnetic tracking system. The solid lines represent the mean difference between the Kinect and goniometer (red) and the Kinect and 3D magnetic system (blue). The dashed lines represent the mean difference ± 95% LOA (1.96*standard deviation of differences).
Table G.1. Test-Retest Reliability.

<table>
<thead>
<tr>
<th>Shoulder Pose</th>
<th>ICC</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>MDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexion to 30°</td>
<td>0.88</td>
<td>22.6°</td>
<td>8.0°</td>
<td>3.0°</td>
<td>4.3°</td>
</tr>
<tr>
<td>Flexion to 60°</td>
<td>0.88</td>
<td>53.4°</td>
<td>5.6°</td>
<td>2.1°</td>
<td>2.9°</td>
</tr>
<tr>
<td>Flexion to 90°</td>
<td>0.21</td>
<td>82.7°</td>
<td>5.9°</td>
<td>2.7°</td>
<td>3.8°</td>
</tr>
<tr>
<td>Flexion to 120°</td>
<td>0.71</td>
<td>112.9°</td>
<td>4.5°</td>
<td>2.2°</td>
<td>3.1°</td>
</tr>
<tr>
<td>Abduction to 30°</td>
<td>0.52</td>
<td>39.3°</td>
<td>5.8°</td>
<td>3.3°</td>
<td>4.6°</td>
</tr>
<tr>
<td>Abduction to 60°</td>
<td>0.90</td>
<td>65.5°</td>
<td>6.5°</td>
<td>1.9°</td>
<td>2.7°</td>
</tr>
<tr>
<td>Abduction to 90°</td>
<td>0.51</td>
<td>91.4°</td>
<td>6.1°</td>
<td>3.6°</td>
<td>5.0°</td>
</tr>
<tr>
<td>Abduction to 120°</td>
<td>0.93</td>
<td>118.6°</td>
<td>7.8°</td>
<td>1.9°</td>
<td>2.7°</td>
</tr>
<tr>
<td>0° External Rotation at 90° Abduction</td>
<td>0.03</td>
<td>9.1°</td>
<td>8.0°</td>
<td>5.6°</td>
<td>7.9°</td>
</tr>
<tr>
<td>30° External Rotation at 90° Abduction</td>
<td>0.80</td>
<td>17.4°</td>
<td>6.4°</td>
<td>2.6°</td>
<td>3.6°</td>
</tr>
<tr>
<td>60° External Rotation at 90° Abduction</td>
<td>0.51</td>
<td>42.1°</td>
<td>3.5°</td>
<td>2.0°</td>
<td>2.8°</td>
</tr>
</tbody>
</table>

ICC = Intraclass correlation coefficient; SD = standard deviation; SEM = standard error of the measure; MDC = minimal detectable change

Table G.2. 95% Limits of Agreement

<table>
<thead>
<tr>
<th>Pose Type</th>
<th>95% LOA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With Goniometer</td>
</tr>
<tr>
<td>Flexion</td>
<td>±11.72°</td>
</tr>
<tr>
<td>Abduction</td>
<td>±11.96°</td>
</tr>
<tr>
<td>External Rotation at 90° Abduction</td>
<td>±25.32°</td>
</tr>
</tbody>
</table>