ABSTRACT

Thesis Title: **A NEW APPROACH TO ASSESS HIGH LEVEL PLANNING UNDERLYING COGNITIVE-MOTOR PERFORMANCE DURING COMPLEX ACTION SEQUENCES**

Theresa C. Hauge, Master of Arts, 2018

Thesis directed by: Assistant Professor Rodolphe Gentili, PhD
Department of Kinesiology

While much work has examined low-level sensorimotor planning, only limited efforts have studied high-level motor planning processes underlying the cognitive-motor performance of complex action sequences. Such sequences can generally be successfully executed in a flexible manner and typically involve few constraints. In particular, no past study has examined the concurrent changes of high-level motor plans along with those of mental workload and confidence during practice of a novel complex action sequence. To address this gap, first a computational approach providing markers capturing performance dynamics of action sequences during practice had to be developed since past relevant works only employed fairly rough metrics. Such an approach should provide concise performance markers (e.g., distances, scalar) while still capturing accurately the changes of structure of high-level motor plans during the acquisition of novel complex action sequences. Thus, by adapting the Levenshtein distance (LD) and its operators to the motor domain, a computational approach was first proposed to assess in detail action sequences during an imitation practice task executed by various performers (humans, a humanoid robot) and with flexible success criteria. The results revealed that this approach i) could support accurately comparing the high-level plans generated between performers; ii) provides performance markers (LD, insertion operator) able to differentiate optimal (using a minimum of actions) from suboptimal (using more than a minimum of actions but still reaching the task goal) sequences; and iii) gives evidenced that the deletion operator is a marker of action sequence failure. This computational approach was then deployed to examine during practice the concurrent changes in high-level motor plans underlying action sequence execution with modulation of mental workload and an individual's confidence in performing the task. The results revealed that as individuals practiced, performance improved (reduction of LD, insertion/substitution and movement time) while the level of mental workload and confidence decreased and increased, respectively. Also, by late practice the sequences were still suboptimal while being executed faster, possibly suggesting different dynamics between the generation of high-level motor plans and their execution. Overall, this work complements prior efforts to assess complex action sequences executed by humans and humanoid robots in the context of cognitive-motor practice, and it has the potential to inform not only human cognitive-motor mechanisms, but also human-robots interactions.

A NEW APPROACH TO ASSESS HIGH LEVEL PLANNING UNDERLYING COGNITIVE-MOTOR PERFORMANCE DURING COMPLEX ACTION SEQUENCES

By

Theresa C. Hauge

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Arts
2018

Advisory Committee:
Assistant Professor Rodolphe J. Gentili, Chair
Professor James A. Reggia
Professor Bradley D. Hatfield
Dr. Garrett Katz

Table of Contents

I. General Introduction

I.a. Motor Control and Learning Processes in Simple and Complex Actions

I.a.1. Hierarchical motor planning and action sequences.

In the last two decades, a growing number of research has challenged the idea that motor planning is a uniform process, with intractable and inseparable elements that stood apart from other cognitive functions; instead, motor planning is a complex hierarchical mechanism (Sober & Sabes, 2003; Sober & Sabes, 2005). Furthermore, research has identified the importance of delineating this hierarchy into categories of ascending complexity and organization to understand the graded nature of human motor control and planning. The initiation of each level of motor planning is related to the challenges associated with the physical space and the intricacy of the motor task. Such a planning process can be divided into first and higher (e.g., second, third or higher) order levels.

The first level or first-order refers to motor planning (often called motion planning) which includes processes that adjust to immediate task demands and activities which require planning to reach a single target or object. For instance, within the context of object manipulation behavior, first-order motor plans process adjustments to immediate task demands such as changing grasp orientation based on the size of the target object or orientation or trajectory of the arm in preparation for a reaching task (Rosenbaum et al., 1993; Rosenbaum et al., 2012; Casartelli et al, 2017).

Higher order (at least second order) motor planning refers to higher-level planning components in which the current behavior is dependent upon tasks (and the goal state of the object) that need to be completed (Fabbri-Destro et al., 2009; Rosenbaum et al., 2012; Wilmut et al., 2013). The level of these higher-order motor planning increases (e.g., second, third, fourth, etc.) as the changes made to a motor plan are the result of a task that is many steps removed from the current plan (Rosenbaum et al., 2012; Meyer, van der Wel, & Hunnius, 2013). During higher-order motor planning, external information can be encoded into the motor plan and thus reorient parameters in the upcoming motor plan (Casartelli et al., 2017). For instance, inversion of the hand when picking up a cup because the end state dictates the object is upside down reflects second-order planning and the prioritization of "end-state comfort" (Rosenbaum et al., 1990). Higher-order motor planning has been examined in both adult populations (Johnson-Frey et al., 2004; Ansuini, 2006) as well as, although modified, in children (Wilmut et al., 2013), toddlers (Chen et al., 2010), and infants (Claxton et al., 2003). This level of planning develops over time as the motor control of the young children develops and may be related to motor reorganization that occurs in children throughout childhood (Thibault & Toussaint, 2010).

Although originally this theoretical framework was presented in a context of object manipulation, it can be extended to more complex skills whose individual actions, as a part of a complex sequence, are influenced by various rules enforced in the environment and whose goal state occurs at a higher level, requiring a number of intermediary steps. Thus, a higher-order motor plan entails reaching a goal state in the course of completing a complex, multi-step sequence of actions that are influenced by other behaviors prior to and following their current state. It must be noted that these complex action sequences include multiple actions which need to follow a certain order to be successfully executed while still providing the flexibility to reach the same goal using multiple strategies. In other words, these flexible movement strategies require a higher level of organization to successfully implement the solution (Naylor & Briggs, 1963). This does not necessarily refer to traditional finger tapping motor sequence tasks, since serial tapping tasks typically engage in sequential movements that are unrelated to each other (that is – they have a lower organization of the component motor actions). More specifically, their motion pattern does not include movements that are dependent upon past and future decisions made to reach a goal state (e.g. Du, Valentini, Whitall, & Clark, 2017). This work will therefore not be addressed in this proposal. One example of a higher-order planning scheme beyond a serial tapping task would be an individual taking necessary steps to prepare and execute an action sequence that needs to be executed in order to solve a task (similarly to a puzzle) while abiding by a set of predetermined rules which limit the problem space (e.g., components that contribute to the process of finding a solution). In this context, an action sequence is an ordered list of motor actions that are executed one after another to achieve a preset goal state within a given environment. Depending upon the rules imposed on the problem space, there is an implicit expectation that a sequence contains necessary actions in an appropriate order. For example, to put together a piece of furniture, there are certain steps that must be completed in sequence to ensure structural stability as one progresses through the instructions. There are some steps that are critically important to complete in a given sequence (high level of organization), while other elements of the complex task can be flexibly maneuvered by the individual to accomplish the task (Naylor & Briggs, 1963).

I.a.2. Examination of lower-order motor planning.

In the field of motor control and learning, lower-order motor planning (also referred to as low-level planning or motion planning) of discrete single movement tasks has been largely investigated via both experimental (e.g., arm reaching movements to single target) and computational (e.g. optimal control; neural modeling) work (Gordon, Ghilardi, Cooper, & Ghez, 2011; Morasso (1981); Soechting & Flanders, 1997, Ashe, 1997; Scott et al., 2001; Sergio & Kalaska, 2003, Gerardin et al., 2000; Hanakawa et al., 2003; Flash & Hogan, 1985, Uno, Kawato, & Suzuki, 1989, Kawato, 1999, Todorov, 2004, Bullock and Grossberg, 1988, Molina-Vilaplana et al. 2006, Gentili et al., 2015a; Gentili et al. 2016). In particular, a large body of literature has examined the motion planning processes to understand the production and selection of effector trajectories with a strong emphasis on linear and angular kinematic analyses. Generally, this work is not done with a sequence of actions, but rather with one atomic action (e.g., center-out arm movement). Most of the previous research in the field of human motor control and learning have explored the cognitive-motor mechanisms, primarily during discrete, well-controlled motor tasks. Typical tasks include planar reaching tasks executed in two (and occasionally in three) dimensions, as well as center-out reaching tasks where individuals must adapt to dynamic and/or kinematic perturbations that distort the trajectory of the effector within the environment (e.g. Sober & Sabes, 2003, 2005; Gentili et al., 2013; Rietschel et al., 2014; Gentili et al., 2015b). This experimental approach allows researchers to control for multiple degrees of freedom in a task, thus providing the opportunity to explore the sensorimotor control and learning mechanisms through detailed kinematic/dynamics analyses of simpler movement tasks. Although altogether behavioral and neuroimaging empirical work as well as computational work have largely examined the low-level sensorimotor planning processes, they only focused to a limited extent on higher-order motor plan.

I.a.3. Examination of higher-order motor planning

There is limited research that has examined higher-order motor planning in humans, and often this work does not involve complex goal-oriented tasks using sequences composed of more than five actions (Rosenbaum et al., 1990; Haggard, 1998). Typically these studies that examined high-level motor planning underlying complex action sequences employ tasks that challenge the performer to engage in high-level motor planning to perform complex action sequences. Tasks employed in these studies include, but not limited to, complex action sequences such as the classic Tower of Hanoi (ToH) puzzle which is the well-known mathematical logic task composed of three pegs and rings stacked in descending diameter in a tower on one peg. The goal is to move the tower of rings from one peg to another while obeying a set of rules which limits the orientation of the rings and the number of moves that can be made at one time (Wunsch et al., 2016). Although the cognitive-motor mechanisms underlying the performance of complex action sequences in healthy individuals has been examined, most of this effort has focused on patients with notable deficits in neural structures that are known to contribute to motor planning and executive functions.

Namely, prior work conducted in healthy individuals has examined the role of executive functions such as working memory and inhibitory control in high-level motor planning during the execution of a complex action sequence using the ToH task (Goldman-Rakic, 1987; Welsh, 1991). These early works revealed that, although inhibitory control and working memory had a limited role on improved performance, the motor planning regions that were engaged in the task recruited attentional resources from different brain areas that are traditionally taxed during working memory and inhibition tasks (Welsh, 1999). In addition, this line of work also suggested that complex action sequence performance is affected by the age of the individuals and that varied training conditions enhance the task representations of individuals to adaptive configurations (Borys et al., 1982; Welsh, 1991; Schiff & Vakil, 2015; Vakil & Heled, 2016). Thus, overall the examination of the cognitive-motor processes in complex action sequences in healthy individuals has primarily used them to examine the role of working memory, inhibitory control, and learning strategies.

However, as mentioned earlier the vast majority of research that examines the neurocognitive processes underlying complex action sequences was conducted in clinical populations with known motor deficits. These studies have heavily focused on clinical populations including those with deficits in high-level motor planning in patients with neural impairments such as Parkinson’s disease, autism and attention deficit/hyperactivity disorder (e.g. Saint-Cyr et al., 1988; Paolo et al, 1995; Hughes, 1996; Chan et al., 2004; Goldenberg et al., 1986; Steenbergen & Gordon, 2006; Chan et al., 2010; Taig et al., 2012; Casartelli et al., 2017). These studies have investigated high-level motor planning processes. For instance, work with autistic children revealed that poor performance in high-level planning are linked to deficits in mental representations of states used in problem solving behaviors. This effort has informed the unique contributions of different brain regions to various complex motor activities in patients with neurophysiological disorders (Goldenberg, Wimmer, Auff, & Schnaberth, 1986; Goel & Grafman, 1994; Rinehart et al., 2006). Deficit in high-level planning processes have also been examined in populations with damage to the temporal lobes or thalamic structures associated with declarative memory and procedural learning. Specifically, patients with Parkinson's or Huntington's disease, damage to the neostriatum and prefrontal cortex could not develop heuristic planning strategies necessary for procedural learning (Saint-Cyr et al., 1988). More recently, the role of cognitive functions such as working memory was evidenced during the execution of a complex action sequence in schizophrenic patients (Chan, Wang, Cao, & Chen, 2010). It must be noted that although such clinical work is critical, it is also possible that the cognitive-motor performance observed is biased since it can be confounded by other comorbidities present in these populations.

I.b. Performance assessment during complex action sequences

Beyond such empirical human work involving healthy and impaired human individuals, one important limitation of this human empirical work is that the metric employed were rough (e.g., movement times, number of moves) and thus were unable to provide detailed information of the structure of the high-level motor plan underlying the action sequence via qualitative (e.g., which action were inserted or deleted and where in the sequence) and quantitative (e.g., number of inserted or omitted action) analysis and therefore provide only a limited assessment of the cognitive-motor performance during execution of such tasks (Hinz, Kostov, Kneißl, Sürer, & Danek, 2009; Vakil & Heled, 2016; Fireman, 1996; Palmer, 1996; Schiff & Vakil, 2015; Vakil & Heled, 2016; Tenorth et al 2013). Also, similarly to human-focused efforts, past robotic work has examined low-level sensorimotor plans whereas only a limited effort has examined high-level motor planning involving cognitive planning functions which determine and assess a motor action sequence for achieving a complex action. It must be noted that as far as we know these high-level motor planning studies did not examine how high-level motor plans can dynamically modified throughout practice. This may be due to the lack of a computational approach able to capture details in the structure of action sequences is available. Thus, there is a need to examine these action sequences to captures in detail differences between performers and/or throughout time beyond what previous work has done by proposing a novel computational approach.

I.c. Neurocognitive and psychological processes in performance of complex action sequences

Although motor control and learning literature has examined key aspects of cognitive and psychological mechanisms that contribute to motor performance the examination of mental workload during practice or learning is lacking in particular in a context of complex action sequences. Thus, there is a need to examine cognitive mechanisms such as attentional control, and working memory in connection with motor learning, practice and performance during complex motor tasks. Namely, while the assessment of mental workload, as reflected by the control of attentional resources, was examined during cognitive-motor performance (Wickens et al., 1983; Miller et al. 2011; Gentili et al., 2014) there is relatively limited research that such a notion exists during motor practice and learning (Bosse et al., 2015; Akizuki and Ohashi, 2015; Shuggi et al., 2017a, 2017b; Hu et al., 2015; Ruiz-Rabelo et al., 2015). In particular, previous effort did not examine changes in mental workload throughout practice during the execution of a complex action sequence. Similarly, while many studies have examined the effect of psychological factors such as confidence (self-efficacy; the belief one has in their abilities to complete a task) on human motor performance there is limited amount of work that has examined changes in confidence in a motor practice or learning context (Chiviacowsky et al., 2012, 2014; Cordova and Lepper, 1990; Ste-Marie et al., 2013; Stevens, 2012; Trempe et al. 2012; Ong et al. 2015, 2017; Saemi et al. 2012; Bandura, 1982). In particular, as far as we know, no study has examined the changes in self-confidence during practice of complex action sequences.

Therefore, overall when considering the prior work that examined high-level motor planning, there is a critical need i) to propose a computational approach to examine in detail the structure of the high-level motor plan underlying the performance of complex action sequence as well as ii) to deploy such a computational approach to examine the concurrent changes in the structure of high-level motor plan underlying performance and the modulation of mental workload and self-confidence during the execution of complex action sequences. To examine these two problems, the proposed work presents two studies. In the first study, a novel computational approach is proposed to assess high-level motor plans during performance of complex action sequences. The results of this study provide a new perspective in the analysis of motor performance that extends beyond kinematic analysis of simple motor tasks. The second study deploys this computational approach within the context of a more constrained task environment to examine the concurrent changes in performance, mental workload and self-confidence during practice of complex action sequences.

II. Study 1: A novel computational approach for complex action sequence performance assessment[1]

II.a. Introduction

Although kinematic analyses are well-developed in the study of motion planning, there is a limited variety of metrics to study the high-order motor planning mechanisms which underlie the control and learning of complex action sequences. Most prior examination of human high-level motor planning has considered only well-constrained problems where individuals have to plan an action sequence while following specific rules (e.g., the well-known ToH or its variations). Behavioral assessment in these studies was largely based on basic metrics such as the response time to complete a task, the total number of moves, number of backwards or reversal moves, or the number of errors made (Goel & Grafman, 1995; Rinehart et al., 2006; Saint-Cyr et al., 1988; Steenbergen & Gordon, 2006). Such measures are often compared to a motor sequence serving as a reference, which often includes a minimal number of steps. This previous work generally does not provide detailed information about the motor action sequences such as which specific actions were added, removed or switched, at which specific positions, in the sequences, information that is critical to assessing the quality of performance and, in particular, its high-level planning component (Goel & Grafman, 1995; Rinehart et al., 2006; Noyes & Garland, 2003; Welsh & Huizinga, 2005; Anderson, Albert, & Fincham, 2005; Hinz, Kostov, Kneißl, Sürer, & Danek, 2009). Such a restricted performance assessment not only provides limited information about the human neurocognitive processes but also the development and assessment of intelligent robotic controllers and human-robot interactions during the execution of complex tasks (e.g., equipment

[1] This first main section has been submitted to publication as Hauge et al. to the International Journal of social robotics (see reference list).

maintenance (Katz, Huang, Gentili, & Reggia, 2016; Katz, Huang, Hauge, Gentili, & Reggia, 2017); cleaning objects of various shapes, (Langsfeld, Kabir, Kaipa, & Gupta, 2016)).

Interestingly, in a similar manner many robotic studies have assessed performance with relatively simple metrics such as questionnaire-based subjective measures (Koenig, Takayama, & Matarić, 2010; Nikolaidis, Ramakrishnan, Gu, & Shah, 2015; Nikolaidis, Hsu, & Srinivasa, 2017; Paxton, Jonathan, Hundt, Mutlu, & Hager, 2017; Steinfeld et al., 2006; Wang, Pynadath, & Hill, 2015; Zhang, Narayanan, Chakraborti, & Kambhampati, 2015), objective time-based measures (Nikolaidis et al., 2015; 35, Steinfeld et al., 2006; 38, Levine & Williams, 2014; 41, Shah, Wiken, Williams, & Breazeal, 2011), and sometimes coarse-grained measures on sequences of discrete steps, such as total step count (Steinfeld et al., 2006; 38, Butchibabu, Sparano-Huiban, Sonenburg, & Shah, 2016; Freedman & Shlomo, 2017; Salter, Dautenhahn, & Boekhorst, 2006; Shah & Breazeal, 2010). Although interesting, overall these various metrics cannot inform a relatively fine-grained performance comparison between humans and humanoid robots. To our knowledge only one prior study has performed a more refined analysis of individual steps in a high-level motor plan that uses alignment algorithms to compare motor action sequences executed in humanoid robots (Tenorth, Ziegltrum, & Beetz, 2013). However, this effort did not examine the quality of human or humanoid performance in a context where both humans and the humanoid robot have to practice a complex action sequence. Such a detailed comparison of the human and robot performance is critical since it could inform one about whether high-level motor plans generated in both performers differ or not, in particular in unconstrained situations where flexible solutions can be identified for learning to complete the complex task.

Given the above, there is a need to examine the quality of performance in both humans and humanoid robots, during complex multi-action motor sequences. Here we focus on action sequences with few constraints and flexible success criteria (e.g., an action sequence can be executed with only a few rules and in versatile ways while still reaching its goal). Our primary interest is an improved characterization and a better understanding of the generation of high-level motor plans during performance of a real-world complex task with minimum task constraints and independently of the nature of the performer (i.e., humans or synthetic such as a humanoid robot). Indeed, in real-world settings it can often be observed that a particular task represented by an action sequence may not be completed due to some major mistakes, and may be completed in multiple ways while still reaching the same goal. There may also exist one (or more) way(s) to execute the task with a minimum number of actions (the latter often characterizing expert performers). A classic example is the case where a manual task is shown to a performer who has to imitate it. In this case, the imitation can be i) a perfect replica of the demonstrated action sequence, ii) different from the demonstration but still reaching the goal, or iii) a failed attempt to reach the goal.

In this current work we take some first steps in examining in detail complex action sequence performance in both a humanoid robot and humans who have to practice a complex task from demonstration provided by a human individual with only few trials. We do this by considering situations where there exists only one unique minimal action sequence to reach the goal but an infinite number of action sequences can be generated and either perfectly imitate the minimal sequence or differ from it while still successfully reaching the goal.

In examining the high-level structure of complex motor sequences, can we propose a computational approach which can determine the functional reasons which lead to: i) performance failure of complex action sequences (which action operations drove this failure) and ii) performance differences between successful action sequences that are minimal (i.e., having a minimum number of steps) and those that are not? Elements of answers to these questions can be provided by detailed examinations of complex multi-action sequence tasks, something that is a challenging problem. Interestingly, other scientific fields, such as DNA sequence analysis in biology or phoneme sequence analysis in linguistics, encounter similar issues. Proposed solutions for sequence comparisons in these fields are generally based on the idea that the number of additions, deletions and substitutions of the bases needed to transform one sequence into another that serves as a reference are computed; a larger number of alterations suggesting that both sequences further differ (e.g., Ho, Oh, & Kim, 2017).

Taking inspiration from these other fields, here we examine how adding, omitting or replacing actions in a motor sequence can affect the imitation learning task performance outcome in terms of failure as well as success where the demonstrated action sequence, having a minimal number of actions, serves as a reference. Such a quantitative analysis is also complemented with a more qualitative examination through graphical representations (e.g., sequence alignment display) to further illustrate any alterations of produced motor sequences by humans and the humanoid robot compared to the demonstration sequence serving as the sequence of reference. In particular, it may be possible that the addition of extraneous actions even between the critical elements of the sequence of reference increases the imitated action sequence length while likely preserving the order of those critical components (the task may still be correctly performed and its goal reached) which may not necessarily be a major determinant of failure to complete the action sequence. However, it is conceivable that omitting or replacing an action of the demonstrated sequence, unless compensated, may represent a greater risk to a successful completion of the action sequence when performing a complex task. If these assumptions are correct, we predict that the occurrence of action omission and/or replacement should be more prominent during failure and should be minimal or null when successfully executing (with a minimum of steps or not) complex action sequences. The results should be consistently observed both through qualitative (e.g., sequence alignment display) and quantitative (e.g., frequency of action omission or addition) analyses. Such an approach should allow for direct comparison of the high-level motor plans generated between performers (humans, synthetic) and/or throughout time during execution of a complex task.

II.b Material and Methods

II.b.1. Participants.

In this study twelve participants (five men and seven women; 18 - 35 years of age) free of any neurological conditions and with normal or corrected-to-normal vision were recruited from the School of Public Health at the University of Maryland, College Park as well as the greater College Park, Maryland area. The study followed the IRB guidelines from the University of Maryland, College Park for which all participants provided their written consent.

II.b.2. Experimental set-up.

The complex task that the human individuals and a humanoid robot (Baxter; Rethink Robotics™) had to perform involved a mock-up hard drive docking station. This station (Figure 1A) was composed of a drawer that, when opened, allowed the participants to manipulate disks in four hard drive slots, each being associated with a LED indicator and a toggle switch, which had to be switched off before inserting or removing the corresponding drive. A red LED associated with each drive indicated a faulty disk that had to be replaced with a new one. If a red LED was initially on and its corresponding switch was pressed, then this LED turned off, indicating that this drive was now disengaged and thus could be replaced. When a disk is successfully replaced and its corresponding switch is pressed, then its green LED turns on indicating that the corresponding disk is now engaged and functioning properly. The custom-made mock-up was controlled by an Arduino processor.

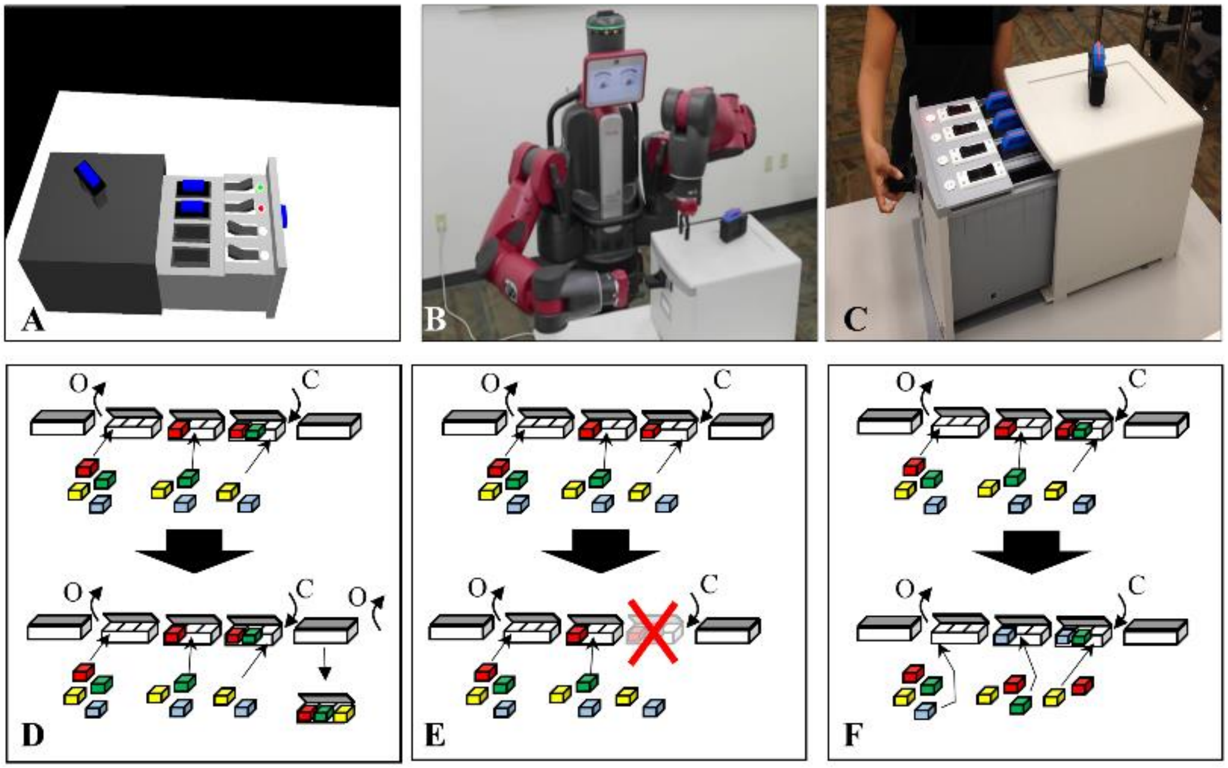

Figure 1. A. Screenshot of SMILE visual software. B. Screenshot of Baxter (Rethink Robotics™ performing the dock maintenance task. C. Image of human subject performing dock maintenance task. Bottom row display the insertion (D), deletion (E), and substitution (F) in a motor task.

II.b.3. Imitation-learning task.

The task considered here differs from more well-known and controlled complex tasks (e.g., Tower of Hanoi) in its inherent complexity, its flexible successful completion criteria, and the limited constraints for the performer to complete a real-world task in a very versatile manner. It was previously employed in cognitive robotic research to examine high-level plan generation in humanoid robots (Katz et al., 2016; Katz et al., 2017), and is well-suited for examining action sequences deployed by the various performers (humans and a humanoid robot) during action sequence imitation. It uses a given demonstration as an action sequence of *reference* having a minimal number of actions.

A humanoid robot (Baxter; Rethink Robotics™) and the human participants had to learn[2] to perform a complex task which consisted of operations on the hard disk drive docking station. Specifically, the participants had to learn to discard a faulty hard disk which was indicated by a red LED being on and replace it with a new one. A human individual demonstrated the action sequence to the performer (humans or robot) through an artificial environment (SMILE; (Huang, Katz, Langsfeld, Gentili, & Reggia, 2015; Huang, Katz, Langsfeld, Oh, Gentili, Reggia, 2015)) which shows the sequence of actions needed to perform the task but does not show any visible force or actor moving the objects. For consistency, this human demonstration was recorded via video and shown to all performers. A screenshot from SMILE is shown in Figure 1A. By having a virtual/invisible demonstrator, a demonstration does not dictate specific arm selection or trajectories for individual actions, and (for the robot participants) greatly simplifies learning. After observing the demonstration, the performers (humans or robot) are required to imitate the task immediately while employing the physical mock-up hard drive station placed in front of them (Figure 1B and 1C). The LED indicating which drive is faulty was constrained to be the same in both the video and the physical mock-up device.

Specifically, via the SMILE virtual environment operated by a human-controlled interface, the video provides a demonstration of this complex action sequence composed of eight atomic actions executed in the following order: 1) the drawer had to be pulled to open the dock to make available the four hard drive slots, their corresponding toggle switches and LEDs (abbreviated dro, for drawer opened); 2) the faulty drive (here the first drive), which could be identified by its red

---

[2] For consistency between humans and the humanoid robot, here the term *learning* is employed in a general manner and reflects performance during the practice throughout the trials [50].

LED, had to be turned off by pressing the corresponding toggle switch (i.e., t1, first toggle); 3) the hard drive had to be picked up to be removed from its slot (ud1), and 4) discarded in a bin (db); 5) a new spare drive was picked up (usp) and then 6) inserted in the empty first slot (ds1); 7) the same toggle switch was pressed again (t1) which turned on the green LED and 8) the drawer was finally closed (drc). The atomic actions in this demonstration sequence are summarized in Table 1. For human-readability we have chosen strings like 'dro' or 't1' that contain more than one letter, but each atomic action is thought of as a *single symbol*, and therefore the smallest unit of activity in this task. In the rest of this paper this particular demonstrated sequence is called the *demonstrated sequence* or *reference sequence*.

Table 1. Orderly sequence of actions forming the complex task demonstrated via SMILE that the humans and the robot have to imitate.

Atomic Action	Abbreviation
Open drawer	'dro'
Press toggle 1	't1'
Pick up drive 1	'ud1'
Discard in bin	'db'
Pick up spare drive	'usp'
Put down in slot 1	'ds1'
Press toggle 1	't1'
Close drawer	'drc'

While the human participants in this study practiced this task by watching a SMILE video, the humanoid robot learned it directly from the SMILE events record that was used to generate the video (Huang et al., 2015). The robot learns by inferring the intentions of the demonstrator rather than mimicking the demonstrator's motor trajectories (which are not provided) based on the algorithm employed in (Katz et al., 2017). After learning, the robot produces a motor sequence that successfully imitates the complex action sequence and that is stored for further analysis.

The human participants practiced in performing the same imitation task. First, before watching any demonstration, an acclimation stage allowed the participants to become familiar with the physical disk drive dock station. A prompt was provided to present the instructions including the meaning of the red and green LEDs and, importantly, also indicating that the LED should be turned off when adding or removing the hard drives. This familiarization stage provided the participants the opportunity to manipulate the physical system before learning the imitation task to avoid any possible misunderstanding of the equipment as confounding factors. This was also done for consistency with Baxter™ which was already trained to manipulate the physical disk drive dock station. Once this familiarization phase was complete, the video was shown, demonstrating the action sequence to the human participants who had to learn to imitate it. Five trials were performed by each human participant, each having the following structure and using the same video. At the start of each trial, the instructions were provided on a computer monitor via the prompt (and also verbally). Once the demonstration was over, the participants had to imitate the action sequence previously observed. As the participants perform the task, a video recording using a digital camera was employed to collect the action sequences they performed, and these actions were subsequently manually transcribed into a symbolic sequence using the symbols in

Table 2. Atomic actions forming the alphabet for disk drive maintenance task.

Atomic Action	**Abbreviation**
Open, close drawer	dro, drc
Press toggle 1, 2, 3 and 4	t1, t2, t3 and t4
Pick up drive 1, 2, 3 and 4	ud1,ud2,ud3 and ud4
Put down in slot 1, 2, 3 and 4	ds1, ds2, ds3 and ds4
Pick up spare drive	usp
Discard in bin	db
Object from left to right hand	ltr

Table 2. At the end of each trial, the hard drive dock station was reset to its initial state and the demonstration in the video ready to be played again in preparation for the next trial.

Success in this task required both: i) the completion of the goal, which was to replace the faulty hard drive with a spare one placed on the top of the docking station, as well as ii) following the basic principle that the LED should be turned off when hard drives are added or removed. If these two criteria were not met, a motor sequence was classified as a failure. It is critical to note that the combination of providing only two success criteria combined with a minimum number of instructions was chosen to minimize the constraints on action performance. As a consequence, it was possible for the participants to successfully perform the demonstrated action sequence without having to exactly imitate the demonstrated motor sequence (which had the minimum number of eight atomic actions executed in the specific order < dro, t1, ud1, db, usp, ds1, t1, drc >; see Table 2 for the complete set of abbreviations used) but also by performing the task in various manners while still reaching the goal (i.e., replacing the drive next to an initially red LED), and respecting the basic rule that LEDs should be switched off while inserting or removing drives. This approach was chosen to promote real-world flexibility in executing the action sequence to capture various high-level plans (i.e., motor sequences) deployed by the humans and the humanoid robot to complete the proposed cognitive-motor task.

II.b.4 Data modeling.

II.b.4.i. The Levenshtein distance applied to complex motor actions.

To assess the quality of performance between any two given action sequences, we adopted a computational approach based on the Levenshtein Distance (LD). LD assesses the similarity between two sequences of symbols (Ho et al., 2017; 47, Levenshtein, 1966; Guo et al., 2013; Navarro, 2001). It has previously been employed in a variety of linguistic analyses as well as for comparing molecular structure in biochemistry (Heeringa, 2004; Leusch, Ueffing, & Ney, 2003). In the field of natural language processing, the atomic symbols are typically letters or phonemes, words are formed by sequences of these atomic symbols, and the LD quantifies how much two words are different. Although several variations of the LD are conceivable, the basic idea is to compute a distance between two sequences by computing the number of changes that need to be applied to the elements of one given sequence to fully match another.

The LD can be applied to the domain of complex motor action sequences by appropriately defining the corresponding alphabet of symbols, sequences, and operations. Thus each atomic "symbol" is here defined as an elementary action that is part of an action sequence. To illustrate this idea, consider a simplified instance of placing three cubes having different colors into a box having three slots. Here, the placement/extraction of a colored cube in/from a specific slot of the box as well as opening/closing the box would be considered to be atomic actions. In this example, the *alphabet* which contains every possible atomic action could be defined as the set of ten symbols {Open_box, Close_box, Place_red_cube_slot1, Extract_red_cube_slot1, Place_blue_cube_slot1, Extract_blue_cube_slot1, Place_green_cube_slot2, Extract_green_cube_slot2, Place_yellow_cube_slot3, Extract_yellow_cube_slot3}[3]. In general, the performance of a complex action sequence can be associated with an alphabet that can be defined as a finite set {A_1, A_2, …, A_i, A_{n-1}, A_n} where A_j is the j^{th} atomic action j among all N possible actions (in the example above $N = 10$). A *motor sequence* would be defined as a finite, ordered list of zero or more actions from the alphabet, potentially with repeats. Using the same example, <Open_box, Place_red_cube_slot1, Close_box, Open_box, Place_green_cube_slot2, Place_yellow_cube_slot3, Close_box> or <Open_box, Place_yellow_cube_slot3, Close_box> are two possible action sequences. The second action sequence indicates that the task consists of three atomic actions: opening the box, as well as placing the yellow cube in the third slot, and closing the box. A sequence of actions in the cognitive-motor domain is analogous to a sequence of letters that form a word (e.g., < d, e, a, r, s >) in the natural language domain or a sequence of nucleotides in a DNA molecule.

In a motor context, the computation of the LD can be associated with three possible operators: i) *insertion* of one action, ii) *deletion* of one action, and iii) *substitution* of one action for another one; the bottom row of Figure 1 provides examples of these operations. These three operations can be employed to transform any motor sequence into another one, by modifying only one action at a time. The *insertion* operator inserts a new action at any position in the action sequence, increasing the length of the motor sequence by one (Figure 1D). The *deletion* operator removes an action at any position in the action sequence, decreasing the length of the sequence by one (Figure 1E). The *substitution* operator replaces an existing action with a new action at the same

[3] Rules for this task should also include preconditions on cube movements. For instance, a cube cannot be extracted from a given slot if there is no cube in that slot. For clarity in this simple example such rules are not introduced. For illustrative purpose this example was kept simple but a more complex alphabet could be generated.

position in the sequence, leaving thus the sequence length unchanged (Figure 1F)[4]. In a motor context, the LD measures the overall distance between two different motor sequences, which is defined as the minimum number of atomic operations required to transform one sequence into the other. The computation of the LD can be efficiently performed by well-established dynamic programming methods such as the Wagner-Fischer algorithm.

II.b.4.ii. Modeling the alphabet of the executed action sequences.

To capture the various ways of successfully completing the disk drive dock task, the set of all potential actions that can be executed (i.e., the alphabet) had to be defined (see Table 2). The hard drive docking station employed here has four hard drive slots, each with one toggle switch and LED. For a given drive there were four actions (one per slot) for picking up a disk, four for inserting a disk, and four for pressing the toggle switch (i.e., total of 12 actions). In addition, the opening and closing of the drawer as well as the pick up of the spare disk and the discard of the faulty drive in the bin represented four additional actions. We also included a possible action ltr ("left to right") for transferring an object between hands because it was observed that the robot as well as certain individuals occasionally transferred an object from their left to their right hand, giving an alphabet consisting of 17 possible actions listed in Table 2.

[4] For consistency with prior work the standard LD considering only these three operators was employed. For this particular study only a very few transpositions and movements were visually identified (Figure 3), so counting them as insertions/deletions had minimal impact on our results.

II.b.5. Action sequence performance analysis.

For the human participants, the video recording resulted in 60 trials (12 participants x 5 trials) which were each analyzed to identify their corresponding motor sequence. In addition, the sequence demonstrated through SMILE (i.e., the sequence of reference) as well as the sequence performed by the humanoid robot were also considered, resulting in a total of 62 action sequences which either led to success or failure based on the criteria defined above. These sequences were given arbitrary identifiers from 1 - 62, except sequence 61 was from the robot and 62 was the demonstration reference sequence.

II.b.5.i. Graphical representation of complex action sequences.

Action sequence alignments were generated to provide a combined visualization of the various imitated action sequences generated by human participants and the humanoid robot with respect to the demonstration sequence (i.e., from SMILE). The action sequence alignments were generated for both the failed and successful (perfect and imperfect) imitated action sequences. This representation allows one to identify where divergences between the motor sequences occur, by aligning all sequences side by side. The spacing between successive actions in each sequence is systematically varied so that the common sub-sequences across all sequences are aligned as well as possible. The alignment allows both a visualization of the generated action sequences at a global level as well as a focus on details in a particular sequence.

II.b.5.ii. Levenshtein distance and operator occurrence analyses.

In order to assess the performance between the performers (humans and humanoid robot), each action sequence imitated by the human or the robot was compared to the demonstration sequence by computing both the corresponding LD as well as the number of occurrences of each of the three operators (i.e., insertions (I), deletions (D) and substitutions (S); where LD = I + D + S). Then, three types of action sequence performance were identified and considered for statistical analyses: i) failed action sequences where either the goal was not reached or it was reached by violating at least one of the execution rules, ii) successful but imperfect action sequences where the goal was successfully reached by employing a different sequence than that demonstrated, and iii) successful perfect action sequences that exactly matched the demonstrated sequence (i.e., those have a LD = 0).

A first analysis was conducted to compare the human and robot action sequences to assess how the human motor plans differed from that employed by the humanoid robot and under which conditions (i.e., failed, successful imperfect/perfect outcome). To this end, the average (computed across trials and participants) LD and the occurrence of its four operators obtained between the action sequence generated by the humanoid robot and those successfully produced by the humans were compared via a series of *t-tests* or *Wilcoxon signed rank* tests depending on whether the assumption of normality was violated or not. In addition, to examine the extent to which insertions, deletions and substitutions are markers of performance success or failure, the average (computed across all trials and participants) LD and the occurrence of its three operators (insertion, deletion and substitution) for the failed action sequences were compared to i) all the successful sequences; ii) the successful imperfect sequences only and iii) the successful perfect sequences only using a reference value of zero (representing a perfect imitation having a zero LD with I = D = S = 0) by employing a series of *t-tests* or *Wilcoxon signed rank* tests depending on whether the assumption of normality was violated or not. Then, to further focus on the action sequences that were successfully imitated, the perfect and imperfect imitated action sequences were examined. To do this, the average (computed across all trials and participants) LD and the occurrence of its three operators for the imperfect action sequences were compared to the value zero (representing a perfect imitation having a zero LD with I = D = S = 0) by employing a series of one sample *t-tests* or *Wilcoxon signed rank* tests depending on whether the data were normally distributed or not. Finally, to assess the contribution of the three operators (insertion, deletion, substitution) to the LD and to determine whether some were employed more frequently than others for all action sequences (failed, imperfect/perfect successful), the LD and its operator occurrences were subjected to a series of paired *t-tests* or *Wilcoxon signed rank* tests based on the normality of the data. These examinations were complemented with a correlational analysis that was conducted to assess the relationships between each of the four metrics and the success or failure of action sequence imitation.

A second analysis aimed to investigate the change in performance throughout the five trial practice. The average values of each of the four metrics plus the number of failed sequences were computed across participants for each single trial and then subjected to a series of one way repeated-measure ANOVA or a Friedman test (with 5 repetitions for the factor trial) based on whether the data followed a normal distribution or not.

A third analysis was applied for the successful and failed imitation action sequences to assess the ability of participants to recall more items correctly at the beginning (i.e., primacy) and/or at the end (i.e., recency) of the action sequences (Deese & Kaufman, 1957; Murdoch, 1962). To do this, the number of insertions, deletions, and substitutions required at each of the eight positions in the demonstrated sequence (SMILE) to transform it into any given human trial sequence were computed. Then, to compare if some operators were employed more frequently than others at the beginning (actions 1 and 2), middle (actions 4 and 5) or end (actions 7 and 8) of the action sequence, the number of insertions, deletions and substitutions were compared for these three periods and each of the eight actions forming the demonstrated sequence by employing paired *t-tests* or *Wilcoxon signed rank* tests based on the normality of the data. This was done for both the failed and imperfect successful action sequences. For all the statistical analyses mentioned above, the false discovery rate (FDR; (Benjamini & Hochberg, 1995; Luck & Gaspelin, 2017)) was employed to control the family-wise error rate for the multiple statistical tests conducted on all the metrics that indexed the humans and humanoid robot performance.

II.c. Results

Analysis of the action sequences revealed that the 62 trials performed included a large variety of sequences among which it is noticeable that the successful sequences (Figure 2; SMILE is the left-most sequence labelled 'T') do not include any deletion or substitution operators contrary to the failed sequences (Figure 3; SMILE is the left-most sequence labeled 'T'; compare also the occurrence of these operators in trials with and without red exclamation mark in Figure 4). Based on the criteria established prior data analysis, 65.57% of the trials were successful (either completely matched the demonstration reference sequence or differed but were still successful) whereas 34.43% of them were not successful in completing the task (Figure 4, see pie chart). The humanoid robot was able to successfully perform the action sequence in one single trial while not strictly following the demonstration with a LD and a number of insertions equal to one and no deletions or substitutions (i.e., LD = I = 1; D = S = 0; see Figure 2 action sequence #12 and in Figure 4 right next to SMILE on the left). An analysis of the 60 human trials revealed that: i) 20 trials (33.33%) were identical to the action reference sequence demonstrated by SMILE (i.e., LD = I = D = S = 0); ii) 19 trials (31.67%) were successful while performing other novel motor sequences (LD = I = 4.053 ± 1.840; D = S = 0) and iii) 21 trials (35.00%) were not successful (LD = 3.190 ± 2.909; I = 2.238 ± 2.879; D = 0.619 ± 0.590; S = 0.333 ± 0.483). In addition, the number of insertions was significantly greater than the number of deletions and substitutions for the successful imperfect action sequences indicating that increased LD was mainly driven by insertions (I vs. D and I vs. S: $z \geq 28$, $p < 0.023$). In contrast, for the failed action sequences the contributions to the LD were more distributed among the three operators while the number of insertions was significantly larger than the number of substitutions (I vs. D: $z = 30$, $p = 0.150$; I vs. S: $z = 43$, $p < 0.023$; D vs. S: $z = 24$, $p = 0.150$).

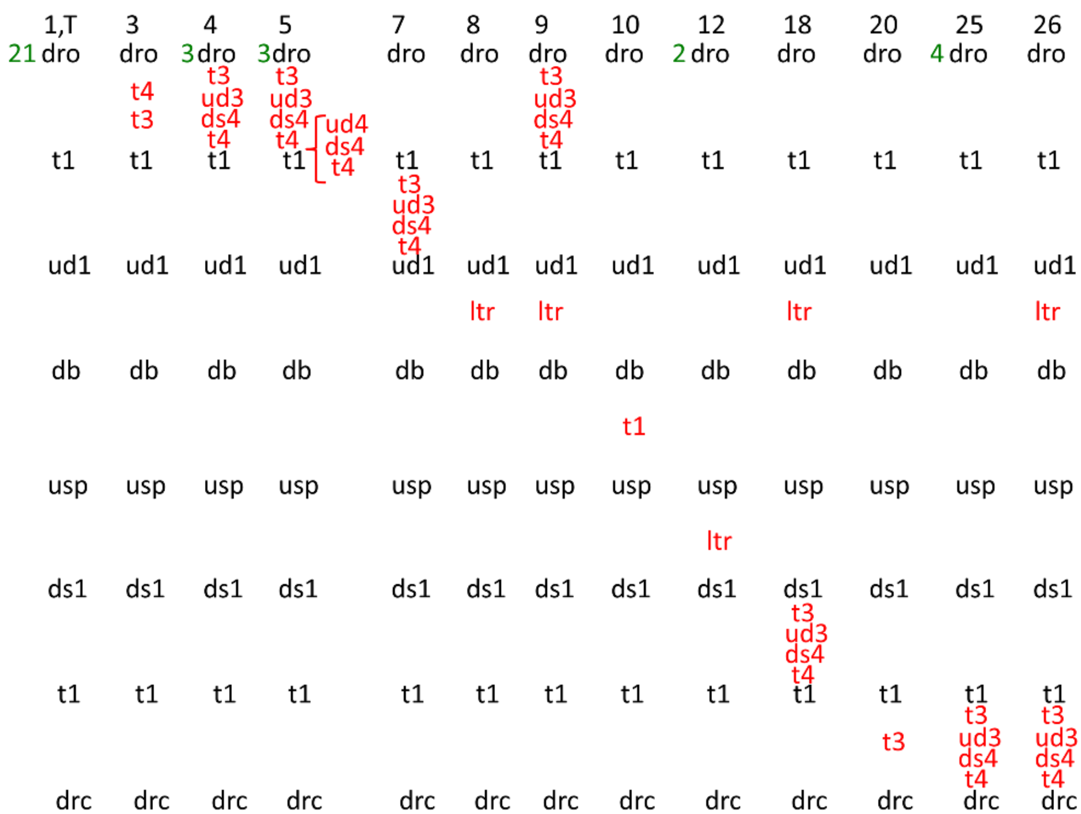

Figure 2. Parallel structure of the various successful motor sequences occurring during human and humanoid robot trials. The leftmost sequence (1, T) is the reference sequence derived from SMILE.

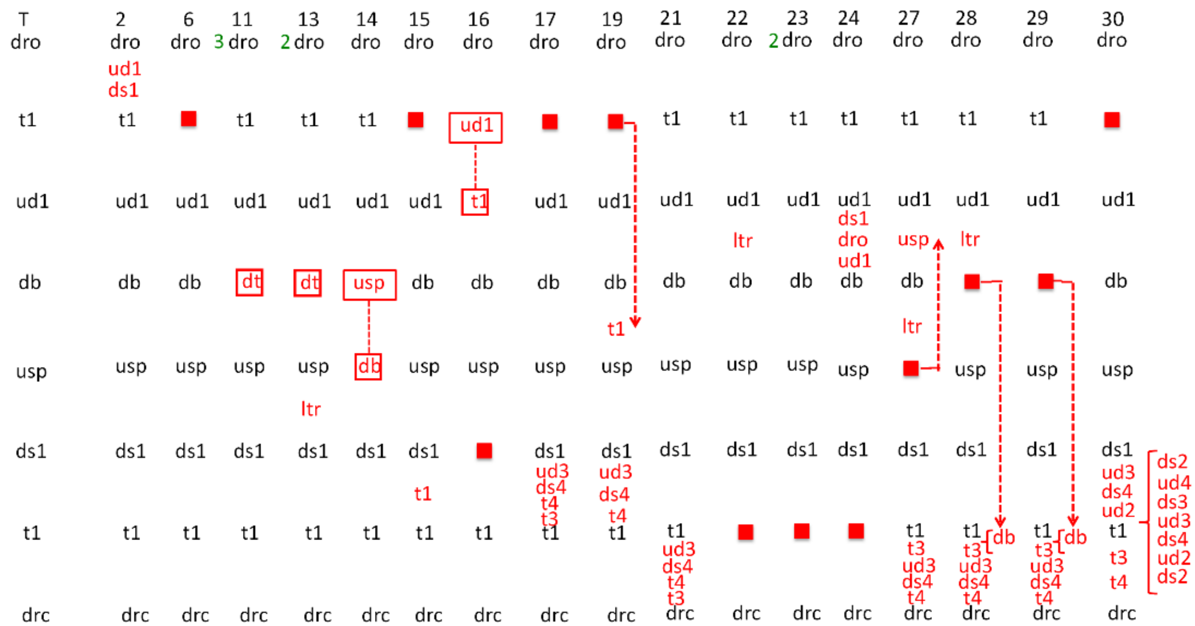

Figure 3. Parallel motor sequence results indicating the presence of deletions (red markers), substitutions (red boxes), or insertions (red actions) in between the reference sequence (SMILE, far left).

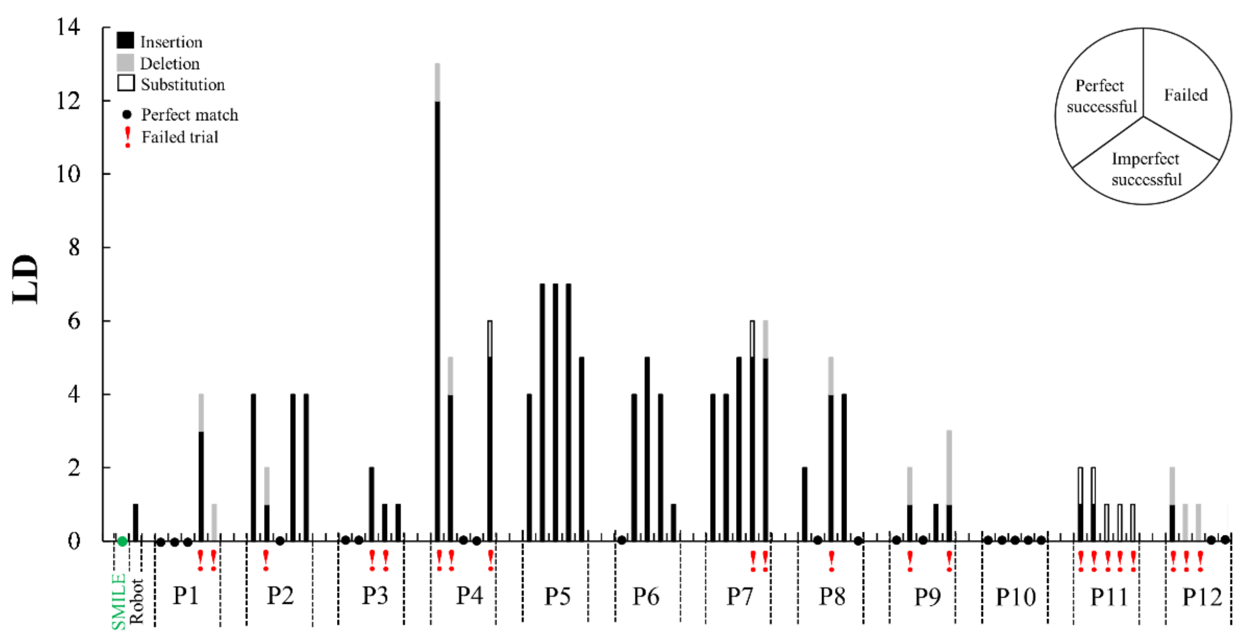

Figure 4. Levenshtein distance (LD) and operator breakdown by participant. Exclamation points indicate a failed trial performance.

III.c.1. Action sequence performance in human individuals and the humanoid robot.

For the failed action sequences executed by human participants, statistical analysis revealed that the average LD and the number of deletions between the reference sequence and the human action sequence were significantly larger than those obtained between the reference sequence and the humanoid robot action sequence (LD: $t(8) = 2.954$, $p < 0.041$; I: $t(8) = 4.438$, $p < 0.013$; compare F and R in Figure 5A and 5C) while the same comparison revealed a similar number of insertions and substitutions ($p > 0.109$ for all comparisons) (see Figure 5B and 5D).

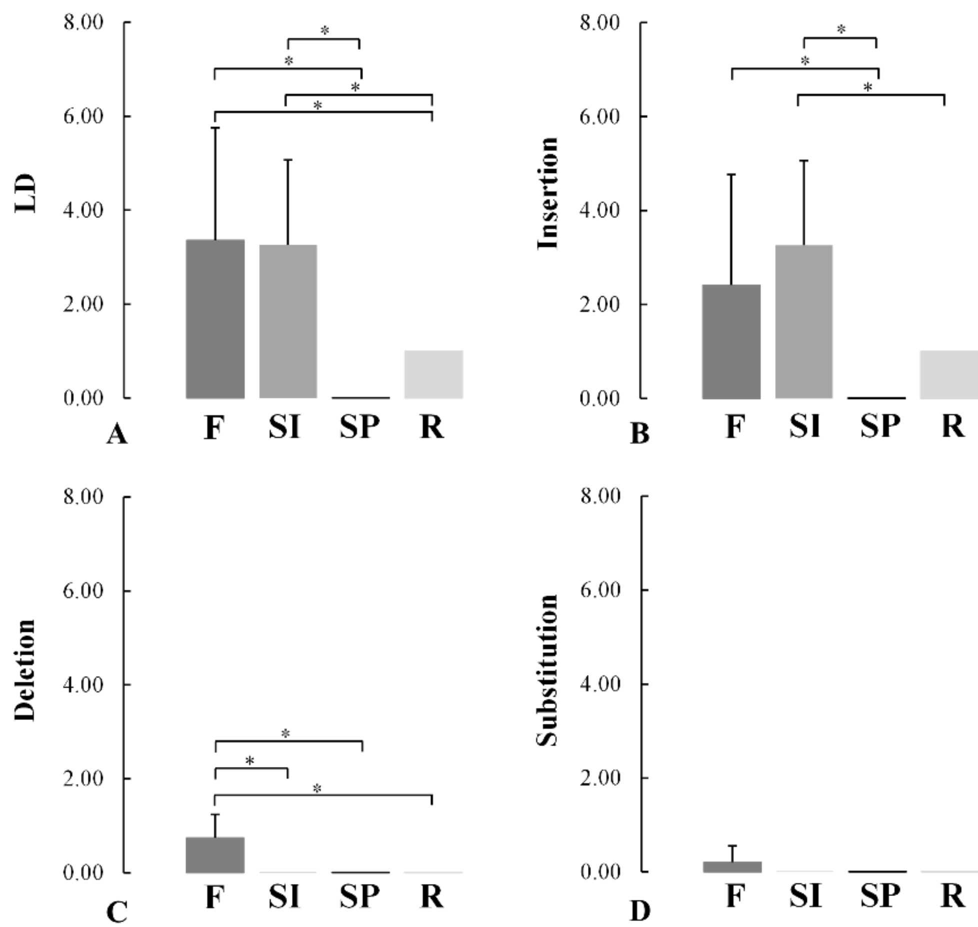

Figure 5. Comparison of overall average LD (A), Insertions, (B), Deletions (C), and Substitutions (D) for failed trials (F), successful imperfect trials (SI), Successful perfect trials (SP), and the humanoid robot trials (R).

The same comparison for the imperfect successful action sequences revealed that the average LD and number of insertions between the reference sequence and the human action sequences was significantly different from that obtained between the reference sequence and the humanoid robot action sequence ($t(6) = 3.316$; $p = 0.040$ for both metrics; compare SI and R in both Figure 5A and 5B).

III.c.2. Failed and successful imitation of action sequences in humans

Statistical analysis revealed that the LD and the number of insertions for the failed action sequences differ from those obtained with the perfect successful sequence (LD: $t(8) = 4.206$, $p = 0.012$; I: $t(8) = 3.079$; $p = 0.015$; compare F and SP in Figure 5A and 5B). In addition, the same analysis revealed that the number of deletions was significantly greater for the failed action sequences compared to those obtained for the successful imperfect or perfect ($t(8) = 4.448$; $p < 0.015$ for both comparisons; compare F and SI as well as F and SP in Figure 5C) action sequences. The same comparison did not reach the significance level for the number of substitutions[5] ($p > 0.120$ all comparisons considered; Figure 5D). Such prominence of deletions in the failed action sequences was confirmed by the correlational analysis that revealed that the number of deletions ($r = 0.681$, $p < 0.001$) was positively and significantly correlated to action sequence failure. This was also observed for the number of substitutions ($r = 0.495$, $p < 0.001$) but not for the number of insertions ($r = 0.219$, $p = 0.090$). The LD was also positively and significantly correlated with action sequence failure, likely due to the contribution of deletion and substitution ($r = 0.519$, $p < 0.001$) (Figure 6A).

[5] The number of deletions and substitutions combined are also significantly larger for the failed sequences compared to the successful imperfect or perfect actions ($p < 0.008$ for both comparisons).

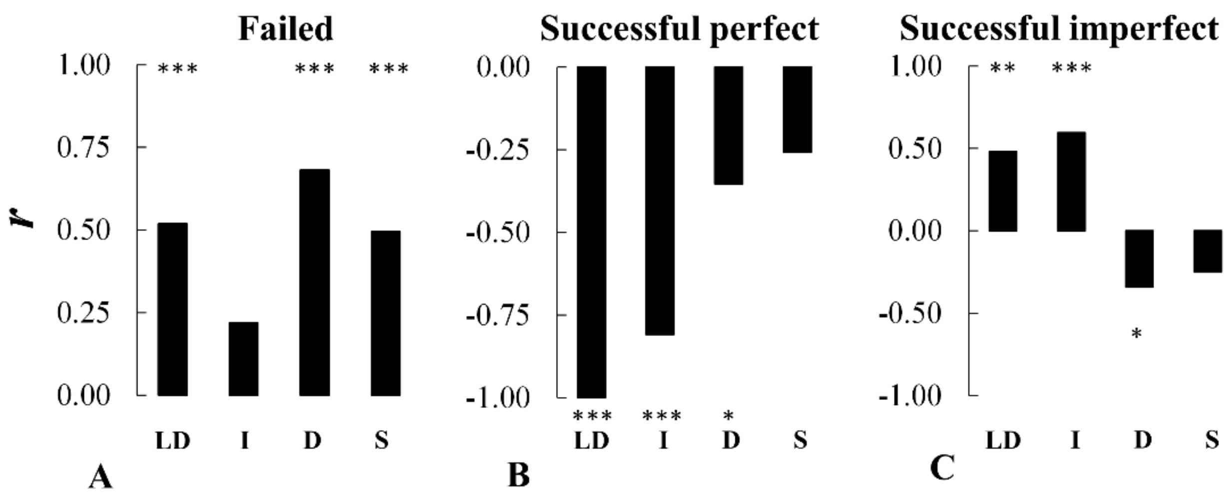

Figure 6. Statistical comparison/correlation between LD and operators for failed (A), successful perfect (SP), and successful imperfect (SI) trials.

III.c.3. Successful perfect and imperfect imitation of action sequences in humans

Statistical analysis also revealed that the LD and the number of insertions (which was the sole contributor to the LD for imperfect sequences) were both greater for the imperfect compared to the perfect action sequences ($t(6) = 3.085$, $p = 0.022$; for both metrics; compare SI and SP in both Figure 5A and 5B). When considering the successful action sequences, about half of the participants generated sequences that had a comparable number of extraneous actions as indicated by similar LDs (i.e., participants #1, 3, 8, 9, 10 and 12).

As expected, the four metrics (LD, number of insertions, deletions and substitutions) were all negatively correlated with the successful perfect action sequences. Significant correlations were obtained for the LD ($r = -1.00$, $p < 0.001$), the number of insertions ($r = -0.809$, $p < 0.001$), and the number of deletions ($r = -0.354$, $p = 0.013$), while only a tendency was detected for the substitutions ($r = -0.257$, $p = 0.056$) (Figure 6B). The same analysis conducted for the successful imperfect action sequences revealed that the LD ($r = 0.481$, $p < 0.001$) and the number of insertions ($r = 0.595$, $p < 0.001$) were positively and significantly correlated with performance outcome.

However, the number of deletions and substitutions were negatively correlated with successful imperfect action sequences, while significant correlations were obtained for the number of deletions ($r = -0.340$, $p = 0.013$) and only a tendency was detected for substitutions ($r = -0.247$, $p = 0.062$) (Figure 6C).

III.c.4. Practice, primacy and recency effects.

Statistical analysis conducted during the practice period revealed that the number of failed trials, the LD, and the occurrence of the three operators remained relatively unchanged throughout practice ($p > 0.05$; all metrics considered). Finally, analysis exploring any recency and primacy effects by examining whether the occurrence of the operators was more pronounced at the beginning, middle or end of the eight actions composing the reference sequence did not reveal any significant differences for the failed or imperfect successful action sequences ($p > 0.05$; all conditions considered). However, a visual examination of the results suggests that for the failed action sequences the deletions tend to occur more at the end and beginning of the sequence while the substitutions tend to rather occur in the middle of the sequence (Figure 7A). For the imperfect successful action sequences, the same visual examination suggested that the insertions tended to further occur at the beginning and end of the sequences (Figure 7B).

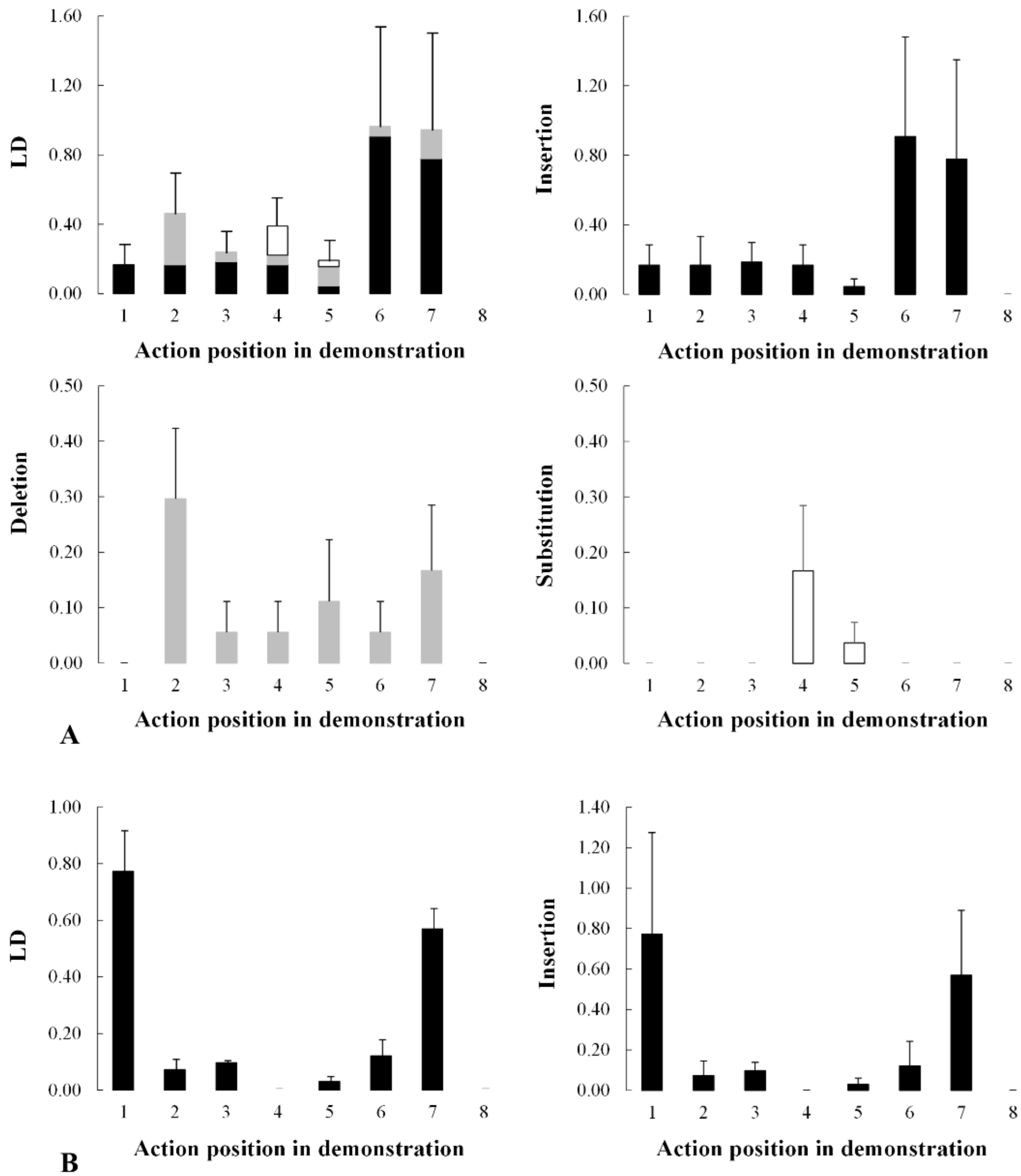

Figure 7. A. LD and operator breakdown for each action in the demonstration sequence for human performers. B. Primacy/recency analysis for the imperfect successful trials analyzing specifically the presence of insertions.

II.d. Discussion

Overall, when examining the humanoid robot and human performance of action sequences under flexible execution conditions like those employed here, the main finding was that, as compared with the robot, human participants generated a wide variety of action sequences not only in terms of success versus failure, but also in terms of the different strategies used in imperfect but successful trials while producing more failed attempts before executing a successful action sequence. Additionally, it was found that LD and insertion counts provide reliable markers able to differentiate perfect and imperfect but successful attempts, whereas deletion was the most sensitive marker of failed action sequences, representing the main contributor to performance failure.

II.d.1. Assessing complex action sequence performance in human and humanoid robots.

By assessing the detailed structure of each action sequence produced by the humans and the humanoid robot it was possible to not only detect if the sequence was a perfect, imperfect (but still reaching the task goal) or a failed imitation of the demonstration, but also to determine specifically what actions were responsible for the performance outcome. Namely, the findings revealed that a third of the human action sequences were not correctly done due to the fact that participants either did not press the release switch to turn off the disk before changing it, or simply failed to reach the task goal (i.e., incorrect order and/or incomplete imitated sequence). As predicted, it appears that the primary operator that led to such failures was the deletion of actions in the sequence. This suggests that this operator primarily contributes to action sequence failure and thus may serve as a robust marker to detect it. A possible explanation is that one cannot recover from deletion of a critical operation of the reference sequence, without reinserting the missing action at the same or a nearby position. Similar considerations may apply to substitution, although to a lesser extent given the lack of statistical significance for that operator when considered alone.

Although no statistically significant findings were identified regarding the distribution of a particular operator among the eight action positions that formed the demonstration sequence, a visual inspection suggests that the number of deletions and substitutions tended to be found at the beginning/end and middle of the failed imitated sequences, respectively. This is somewhat surprising as it is the opposite of primacy and recency effects seen in working memory studies, since the actions placed at the beginning and end (see actions #2 and #7 in Figure 7A) are the most affected by errors. However, direct comparison with prior primacy/recency work is somewhat difficult since these notions were identified in pure task memorization and not complex motor tasks (Postman & Phillips, 1965; Talmi, Grady, Goshen-Gottstein, & Moscovitch, 2005; Morrison, Conway, & Chein, 2014).

Our approach revealed that half of the human action sequences perfectly replicated the demonstrated sequence (i.e., demonstration-human LD = 0) while the other half did not (i.e., demonstration-human LD > 0). These successful but imperfect imitated sequences were of particular interest since they represent different ways to complete the task, something that is often observed in real-world settings. Interestingly, when examining the operators, our method revealed that the lack of perfect match of these successful sequences was exclusively due to the insertion of additional unnecessary actions during sequence imitation (i.e., no deletions or substitutions for the successful sequences were identified). For instance, instead of picking up the disk with the right hand and directly placing it in the available slot, the disk could be grasped with the left and then transferred to the right hand to be finally placed in the slot. Such change of hand was observed in humans and also in the humanoid robot. Adding such extraneous actions (which could be done multiple times) increases the length of the sequence without however compromising the success of the task completion. Thus, while the presence of insertions may prevent a perfect imitation, they may have a low probability to impede successful task completion, contrary to deletions and substitutions. Possibly, the insertion operator was less likely to prevent successful task completions since inserting actions before/after each of the eight actions of the demonstrated sequence does not pose a problem as long as these actions are produced in the correct order while following the rules to complete the task such as that no hard drive could be extracted with the LED turned on.

Since here there was only one optimal sequence which served as a reference (i.e., the demonstration), the action sequences with a zero LD[6] were all identical to the demonstration and thus between themselves. However, those having the same, but non-zero, LD were not necessarily identical to each other since it is unknown to which actions the operators were applied. In particular, action sequences that have the same non-zero LD do not necessarily use the same high-level motor plan since the action sequence could have the same length but include different actions. Therefore, here having the same LD was a necessary but not a sufficient condition to state that the high-level motor plan underlying two sequences was the same. To further investigate high-level motor plans deployed by the human and the humanoid robot for sequences with the same LD, the visual representation of actions sequences allows one to examine if an exact same path captured two or more sequences and which actions drove any discrepancies between those sequences. For instance the action sequence generated by the humanoid robot (see sequence #12 in Figure 1 and Table 3), and another one produced by a human participant (see sequence #10 in Figure 1 which corresponds to trial #5 for participant #3), both have a LD = 1 from the demonstration (only one

[6] Here since only one optimal sequence (minimum number of step) of reference (i.e., demonstration) existed, if two or more sequences have a zero LD thus they are identical. This would be different if several optimal sequences existed since two trials might have non-zero LD with each other, but zero LD with the respective optimal sequences they match.

insertion, but a different action inserted at a different position). Without any visualization the same LD obtained for the two sequences would not reveal their different structure.

Table 3. Participant, trials and outcome for action sequences shown in Figures 2 and 3.

Sequences in Figures 2 and 3	Number of Trials	Participant/Trial Number	Success	Failure
1	21	SMILE; P1(T1,T2,T3); P2(T3); P3(T1,T2); P4(T3,T4); P6(T1); P8(T2,T5); P9(T1,T3); P10(T1,T2,T3,T4,T5); P12(T4,T5)	1	0
2	1	P3(T3)	0	1
3	1	P8(T1)	1	0
4	3	P2(T1,T4,T5)	1	0
5	3	P5(T2,T3,T4)	1	0
6	1	P1(T5)	0	1
7	1	P5(T1)	1	0
8	1	P9(T4)	1	0
9	1	P5(T5)	1	0
10	1	P3(T4)	1	0
11	3	P11(T3,T4,T5)	0	1
12	2	Baxter; P3(T5)	1	0
13	2	P11(T1,T2)	0	1
14	1	P2(T2)	0	1
15	1	P9(T2)	0	1
16	1	P9(T5)	0	1
17	1	P4(T2)	0	1
18	1	P7(T3)	1	0
19	1	P8(T3)	0	1
20	1	P6(T5)	1	0
21	1	P8(T4)	0	1
22	1	P12(T1)	0	1
23	2	P12(T2,T3)	0	1
24	1	P1(T4)	0	1
25	4	P6(T2,T4); P7(T1,T2)	1	0
26	1	P6(T3)	1	0
27	1	P4(T5)	0	1
28	1	P7(T4)	0	1
29	1	P7(T5)	0	1
30	1	P4(T1)	0	1

II.d.2. Comparison to existing approaches to assess complex action sequences

The current results complement past work by extending prior findings from studies conducted in both human motor behavior and robotic domains that evaluated and compared human and/or robot performance during execution of complex tasks (e.g., Towers of Hanoi, equipment maintenance, material cleaning) with relatively basic task-specific metrics (e.g., number of moves, backwards/reversal moves, errors, task completion time). Namely, here we have deployed a computational approach able to provide a relatively detailed examination of the structure of high-level motor plans underlying action sequences via qualitative (visualization) and quantitative (LD and occurrence of specific operators) approaches while also being more generalizable to various complex tasks (Tenorth, Ziegltrum, & Beetz, 2013; 20, Goel & Grafman, 1995; Saint-Cyr et al., 1988; Noyes & Garland, 2003; Welsh & Huizinga, 2005; Anderson et al., 2005; Hinz et al., 2009; Katz et al,. 2017; Langsfeld et al., 2016). More specifically, our current study complements past robotic work that has assessed complex motor tasks by employing relatively coarse-grained metrics (e.g., execution times, success rates (Paul & Shimon, 1979; Omote et al., 1999) or questionnaire-based subjective measures (e.g., trust, mental workload, situational awareness) to assess human-robot teaming (Nathan et al., 2010; Nikolaidis et al., 2015; Nikolaidis et al., 2017; Paxton et al., 2017; Steinfeld et al., 2006; Wang et al., 2015; Zhang et al., 2015). Our work also complements more refined metrics based on time-based measurements (e.g., human idle time, reaction times of the robot/human teammate (Nikolaidis et al., 2015; Steinfeld et al., 2006; Levine & Williams, 2014; Shah et al., 2011) or the segmentation of task execution into action sequences (e.g., number/cost of actions performed, communications exchanged per unit time (Nathan et al., 2010; Steinfeld et al., 2006; Butchibabu et al., 2016; Freedman & Shlomo, 2017; Salter et al., 2006; Shah et al., 2010)). Interestingly, by employing a clustering-based approach Nikolaidis and colleagues have also examined the sensorimotor performance but with a different distance metric and only for low-level motor behavior (Nikolaidis et al., 2015). In their work, action sequences were modeled as Markov processes, and distances were computed between the transition probability matrices. The task involved the robot maintaining its grasp on a single object, and the robotic actions available were to position the object at different locations and orientations. Our distance metric proposed here extends this clustering-based approach to higher-level cognitive tasks, and since it does not rely on a Markov assumption, it can capture longer-term dependencies in action sequences. It must be noted that only a few prior studies employed LD in studying human or robot motor control, however the tasks considered were much simpler compared to that employed here, being more sensorimotor in nature and with a limited or no cognitive component (e.g., rhythmic motor primitive sequences; finger tapping sequences (Gorbenko & Popov, 2015; Holm, Karampela, Ullén, & Madison, 2017).

The application of the LD to the motor domain can account for both the level of abstraction and the physical constraints. Indeed, in the motor domain, the “alphabet” can include not only the action to be selected depending on the cognitive-motor stage using high-level abstract representations (e.g., opening the drawer), but also on sensorimotor coordination for action implementations (e.g., use of the left, right effectors or both) which rely on a lower level of abstraction. While adding performance constraints (e.g., requirements to exactly replicate the demonstration with the right arm only; extra execution rules) can mitigate or even solve this problem, the task becomes less realistic. Also, in the motor domain, humans and/or robots interact with physical systems that provide task constraints that are naturally enforced. For instance, in the hard-drive maintenance task presented here, the dock must be opened first before any manipulation of the disks, which is an implicit rule due to the physics of the system to be handled. Another example is the situation where an imitator and a demonstrator have different physical constraints (e.g., difference in the number of degrees of freedom). However, such specificities may not be encountered for motor tasks that are less complex and with more constraints, limiting the use of different effectors and implicit rules. For instance, the studies mentioned above that employed LD in humans or robots did not face these two problems due to the use of fairly simple tasks (Gorbenko & Popov, 2015; Holm, Karampela, Ullén, & Madison, 2017). As such, the proposed approach can be employed to assess complex action sequences while still complementing the previous proposed simpler metrics to provide a comprehensive assessment of complex task performance by human and humanoid robots.

II.d.3. Applications to human, humanoid robots and human-robot interactions.

Our approach to analyzing human motor sequences could be employed to assess performance in both humans and humanoid robots. Specifically, by understanding the generation and adaptation of high-level motor plans in humans, results from our approach could inform the design of neurocognitive architectures for cognitive robots, enhancing their learning and performance capabilities. Importantly, our approach could also be employed to assess how human and humanoid robot high-level motor plans differ, allowing testing and prediction of the quality of human-robot team dynamics. Namely, it is possible that a humanoid robot learns to perform a complex maintenance task in a certain way whereas humans execute it in a manner which may be very different. Thus, the proposed approach could accurately quantify how much the humanoid robot and humans differ when completing the task. As such, we could predict that as the level of discrepancy between the humanoid robot and human increases, (e.g., elevation of LD and of insertions occurrences) the human-robot team dynamics would translate from an adaptive to a maladaptive state (Losey, McDonald, Battaglia, & O'Malley, 2018; Miller et al., 2013; Miller et al., 2014; Shuggi, Shewokis, Herrmann, & Gentili, 2018). Thus, such an approach focusing on high-level motor planning can complement the previous work that has examined shared control at lower sensorimotor levels for enhancing human-robot interactions (Ishiguro, 2007; Gentili, Oh, Kregling, & Reggia, 2016; Shuggi et al., 2018; Gillespie, 2001; Boehm, Ghasemi, O'Madhrain, Jayakumar, & Gillespie, 2016). Also, such an approach could reveal whether some specific changes in the sequence (e.g., insertion of a specific action) may be the main drivers to promoting adaptive or maladaptive team environments. This could be done in a context of performance where the humanoid robot and humans have been trained separately or are being trained together. In particular, the fact that here the humanoid robot could perform consistently the task learned after one single demonstration whereas this was not observed for humans (likely needing more practice) suggests that the difference in robot and human learning capabilities would likely influence the human-robot collaborative learning dynamics. Also, although imitation learning in humanoid robots and humans was considered here, our approach could be employed in learning without any demonstration and exclusively with robots or humans. Regarding the latter, this work could serve to compare high-level motor plans in both healthy and impaired individuals (e.g., attention deficit hyperactivity disorder) informing thus the compromised underlying cognitive-motor processes. In particular, our approach could inform how the motor plan generation and more generally the human cognitive-motor processes are adapted during performance and learning of new complex tasks.

II.e. Conclusion and future work

This work is just a first step in examining qualitatively and quantitatively the cognitive-motor performance in humans and robots during practice of complex action sequences. Our results provide different but complementary metrics to those already existing to assess complex task performance by quantifying the differences between action sequences. The LD in the cognitive-motor domain can be used with varying levels of abstraction in the context of physical constraints inherent in real-world interactions. While the LD is informative, the occurrence of its specific operators may, to some degree, provide a more refined assessment of cognitive-motor performance. In addition, although this work was conducted in a context of task imitation, it can also be applied to motor performance and learning outside such an imitation framework. However, a limitation of processing the insertions, deletions and substitutions separately is that, when considered independently, each operator count is no longer a formal distance metric, thus limiting the use of multidimensional scaling. Another limitation is the situation where a task includes many trials with many steps. In this case, the visual representation and in particular the action sequence alignment used here may not be able to represent clearly all the trials at once but only under specific configurations. Although interesting this work did not examine changes in performance along with modification of mental workload and confidence while using more standard and controlled, albeit less realistic, complex tasks such as the ToH. Therefore, in the second part of this work the proposed computational work was deployed to assess the performance along with the modulations related to mental workload and self-confidence during practice of action sequences to solve the ToH task.

III. Study 2: Concurrent examination of motor performance, mental workload and confidence during practice of complex action sequences

III.a. Introduction

Cognitive-motor performance depends upon how attentional resources are allocated for a given task and the resultant effects this allocation will have on an individual's mental workload (Wickens et al., 1983; Miller et al. 2011; Gentili et al., 2014). As task difficulty increases, an individual must appropriate more attentional resources toward the task, which increases the mental workload experienced by the individuals. An excessively elevated cognitive workload can impact motor performance (Berka et al., 2007; Miller et al. 2011; Gentili et al., 2014). Despite the wealth of information related to mental workload in a motor performance context, there is relatively limited research that examined mental workload during motor practice and learning (Bosse et al., 2015; Akizuki and Ohashi, 2015; Shuggi at al. 2017a, 2017b). For instance, one study measured mental workload and self-efficacy to assess how feedback influences performance, but no retention or transfer tests were considered that might have predicted the impact of this feedback on mental workload and future performance (Bosse et al., 2015). In another study, Akizuki and Ohashi (2015) considered the principles of the optimal challenge point framework to assess the relationship between a motor learning task and mental workload while manipulating the task difficulty. In the same vein, Shuggi and colleagues (2017a) examined the relationship between the mental workload and performance while practicing a motor task under various levels of task demand practice and revealed that increased task difficulty resulted in elevated demand which beyond a certain level can delay motor performance improvements.

Another study revealed that during intensive surgery procedures, a simulated laparoscopic technique could reduce cognitive workload. However, self-efficacy and retention tests were administered only immediately following the activity and did not include a delayed retention test and/or a transfer test to assess learning beyond practice (Hu et al., 2015; Magill & Anderson, 2014). Similarly, a study by Ruíz-Rabelo and colleagues (2015) examined mental workload and fatigue during laparoscopic tasks. Although interesting, motor learning was not assessed because there were no retention or transfer tests administered after the practice condition (Ruíz-Rabelo et al., 2015). Importantly, these prior works did not examine changes in mental workload and performance dynamics throughout practice during the execution of a complex action sequence. Past work has examined the effect of psychological factors such as confidence of individuals in their abilities to complete the task (Trempe et al. 2012; Ong et al. 2015, 2017; Saemi et al. 2012; Bandura, 1982). While most of this work was conducted in a performance context there is a limited amount of work that has examined changes in confidence during motor practice and learning. For instance, a relationship between motor performance and perceived competence during completion of a motor task was observed (Chiviacowsky et al., 2012). Learning outcomes in a motor skill are influenced by confidence (Chiviacowsky et al., 2012; Chiviacowsky, 2014; Cordova and Lepper, 1990; Ste-Marie et al., 2013), and the difficulty/performance of the task is strongly linked to the concept of confidence in performers (Stevens, 2012). However, as far as we know, no study has examined the changes in confidence during practice of complex action sequences. In addition, no study investigated the concurrent changes in motor performance, mental workload, and confidence during practice of complex action sequences.

Therefore, the aim of this second study was to examine the concurrent changes in the structure of the high-level motor plans underlying performance along with modulation of mental workload and confidence during practice of complex action sequences. We predicted that as practice progress: i) the LD and number of insertions and substitutions should decrease. The number of deletions should be positively related to the number of failed trial (i.e., should decrease if the number of failed trials also decrease during practice); ii) the level of mental workload and confidence should decreases and increase, respectfully. We also predict that i) by late practice the action sequence generated should be optimal or close to optimal (i.e., LD similar to zero) and ii) the deletion should decrease with the decrease of failed action sequences. In addition, an analysis to examine transfer on different level of task difficulty was conducted. Although this analysis was exploratory, the general prediction would be that as the task demand increases: i) the LD, number of insertions, deletions and substitutions should increase; ii) the level of mental workload and confidence should decreases and increase, respectively.

III.b. Methods

III.b.1. Participants

In this study, twenty right-handed participants (male = 9, female = 12, ages 18-35 years) free of any neurological conditions and with normal or corrected-to-normal vision were recruited from the School of Public Health at the University of Maryland, College Park as well as the greater College Park, Maryland area. Exclusion criteria included any neurophysiological disorders or extended experience with mathematical logic and other disentanglement puzzles. The study followed the IRB guidelines from the University of Maryland, College Park for which all participants provided their written consent.

III.b.2. Experimental procedure

III.b.2.i. Task: The Tower of Hanoi

The Tower of Hanoi (ToH), invented by Frenchman Édouard Lucas in the 19th century, is a mathematical logic puzzle with a simple structure that belies the complexity of its problem space. It is composed of at least three disks stacked atop each other in descending order of diameter. These disks are stacked onto a peg, which has two other identical counterparts that are aligned linearly horizontal to the performer. The task is governed by three rules:

a) Only one disk may be moved at a time;
b) A disk may not be placed on the table or held in the hand while another disk is manipulated in space; and
c) A larger disk may not be stacked on top of a smaller disk.

A traditional problem for the ToH involves moving a complete tower from peg A to peg C, while obeying the rules established for the problem space.

The ToH has a recursive solution based on the number of disks present. For h_i = optimal move length to reach a goal state in which i = number of disks, $h_i = 2^i - 1$. (Palmer, 1996). It should be noted that this is true provided that the task begins in one tower and ends in another tower in the goal state; when the initial state is altered such that the disks do not form a complete tower, the number of moves required to reach the goal state is changed. If the initial state represents any transition state (h_t) present in the sequence toward the goal state, the number of moves to the tower-ending goal state $h_t < h_i$.

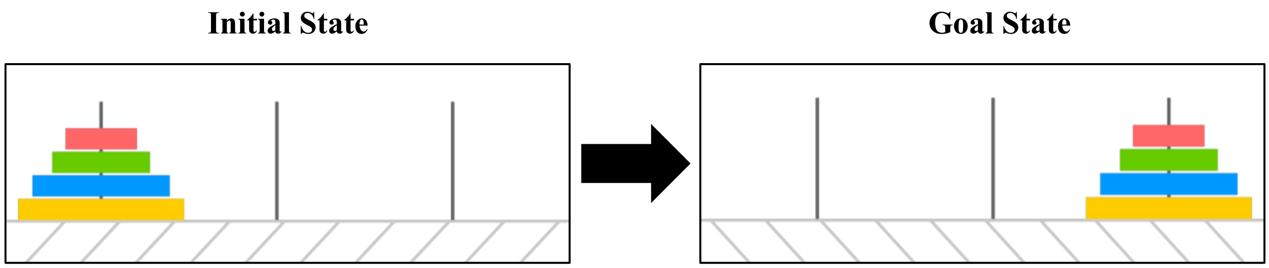

Figure 8. An illustrative example of a conventional Tower of Hanoi problem space. Individuals must move the stack of rings from the leftmost peg to the rightmost peg while abiding by the three rules of play. Photo source: https://www.kirupa.com/html5/tower_of_hanoi_puzzle.htm

It should be noted that h_i can be manipulated based on a number of constraints placed onto the task space, including but not limited to i) the number of disks; ii) the number of pegs; iii) changes to the initial state of the disks; iv) changes to the goal state; and v) adding in additional constraints such as interleaving towers of different towers (e.g. Egan & Greeno, 1974; Palmer, 1996; Welsh, Satterlee-Cartmell, & Stine, 1999). In many instances, task difficulty is determined by the number of moves required to reach the goal state (Welsh et al., 1999). Due to the recursive nature of this problem, the increase in number of rings leads to a nearly exponential increase in time to complete the tower successfully and the difficulty to reach the goal state at peg C.

There is only one optimal solution to each tower goal state in the ToH. Upon inspection of the four-disk ToH task, the optimal path length h is for one optimal path, provided that the puzzle problem presented is in its original form. That is, when the tower is stacked on peg A and the goal state is to move the tower to peg C, there is one way that the disks can be moved that will result in the optimal path $h_4 = 15$. Any deviation from that optimal path is evident in $h_4 \neq 15$. Success is still attainable if the goal state is reached and the rules are not violated, however it would be classified as a suboptimal solution.

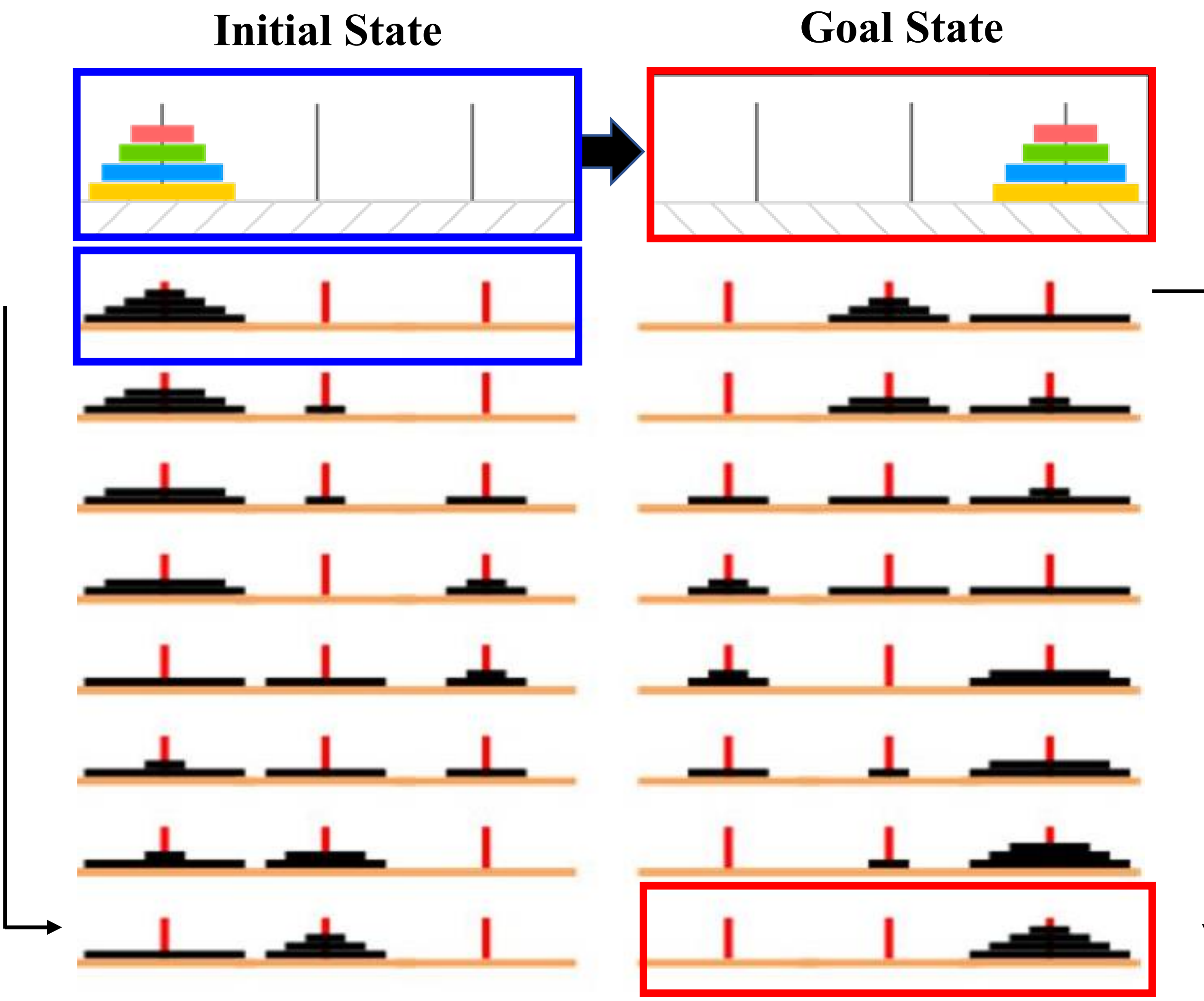

Figure 9. This is a step-by-step breakdown of the optimal solution path for a four-disk tower in the Tower of Hanoi. There is a total of 16 states, punctuated by 15 moves that must be completed in this order. Photo source: http://mathworld.wolfram.com/images/eps-gif/TowersofHanoiSolution_700.gif

III.b.2.ii. Data collection

First, participants completed survey questionnaires confirming their general health, neurological health, and handedness. Consent forms were read and completed by participants prior to introduction to the experimental task. Participants sat at a table where a physical Tower of Hanoi was placed (custom 3D printed, 42 x 12 x 11 cm). A written prompt explaining the task was read and available for viewing before completing the trials. The prompt provided historical background of the task and included instructions to complete the task in the minimum number of moves as possible within the time limit. During the familiarization stage, participants manipulated the physical tower with 6 rings for three minutes, any questions about the rules of the problem space or the instructions on the prompt were answered. Further, the prompt was displayed for the entirety of the session as a reference for the participants. Participants were not informed of the recursive solution strategy during the written/spoken prompt portion of the exercise.

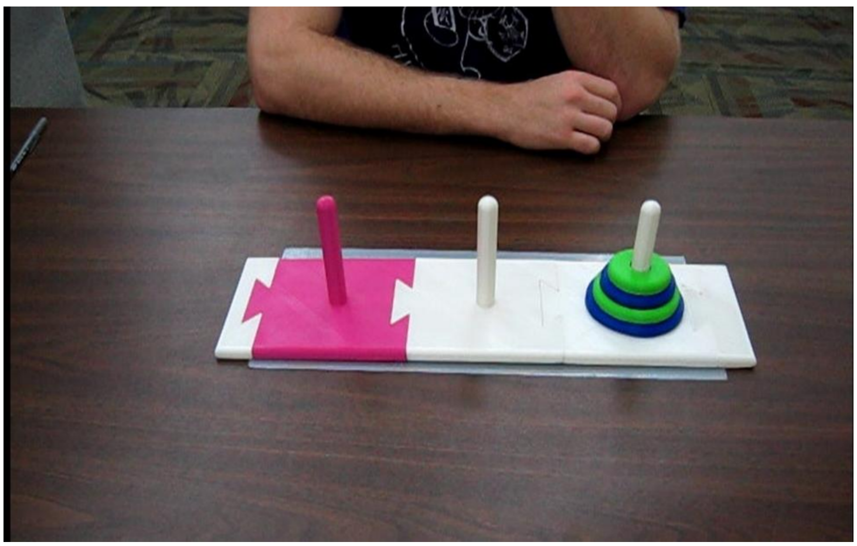

Figure 10. Experimental setup for the ToH task space used in this study. The participant sat behind a cardboard screen prior to the beginning of each trial, and the trial was initiated after the barrier was removed. The goal state of the game is to the right of the participant, which is pink in this model.

After the familiarization phase, the initial state of the task was obscured by a physical barrier. The researcher read additional instructions to the participants, specifying the number of disks on the tower, the time limit for each trial, and the goal state and conditions of the task. Then, the barrier was removed from in front of the 4-ring ToH setup. A digital camera initiated recording for data collection of the trial, and the participant was told that they may initiate the task. The researcher started a timer to track the participant to reach the goal within the time limit.

Participants completed 5 blocks of trials, each consisting of 5 trials. During each trial, researchers observed participants for instances of rule violations and trial time termination. If the participant did not reach the goal state within the time limit, they were informed that all time had elapsed for the trial, and the physical barrier was replaced in front of the participant to reset the task for the next trial. Participants who violated a rule during a trial were informed when the trial was complete which rule they violated during their trial. For the practice blocks of four-ring ToH task, each trial at a time limit of 60 seconds. Completion of a trial occurred either if the time elapsed for the trial or the participant successfully moved the tower to peg C and dropped the final ring atop the stack. The NASA-TLX and confidence surveys were administered after each block of practice.

Once all practice blocks were completed, the participants waited for 10-15 minutes as a rest interval and completed surveys to assess mental demand and task difficulty (NASA TLX) as well as how confident they were in their capability to perform the task (confidence surveys). In order to assess for the effect of transfer an immediate post-test following the practice period consisted of an immediate retention task which consisted of 1 block of 5 trials at the same level of task difficulty that was practiced (i.e., four disks), an easier (i.e., three disks) and a harder (five disks) level of task difficulty relative to practice task difficulty. Post-test conditions were counterbalanced when introduced to participants to minimize any task order effects. The mental demand and confidence were assessed by employing the same surveys previously mentioned after each block of the post-test.

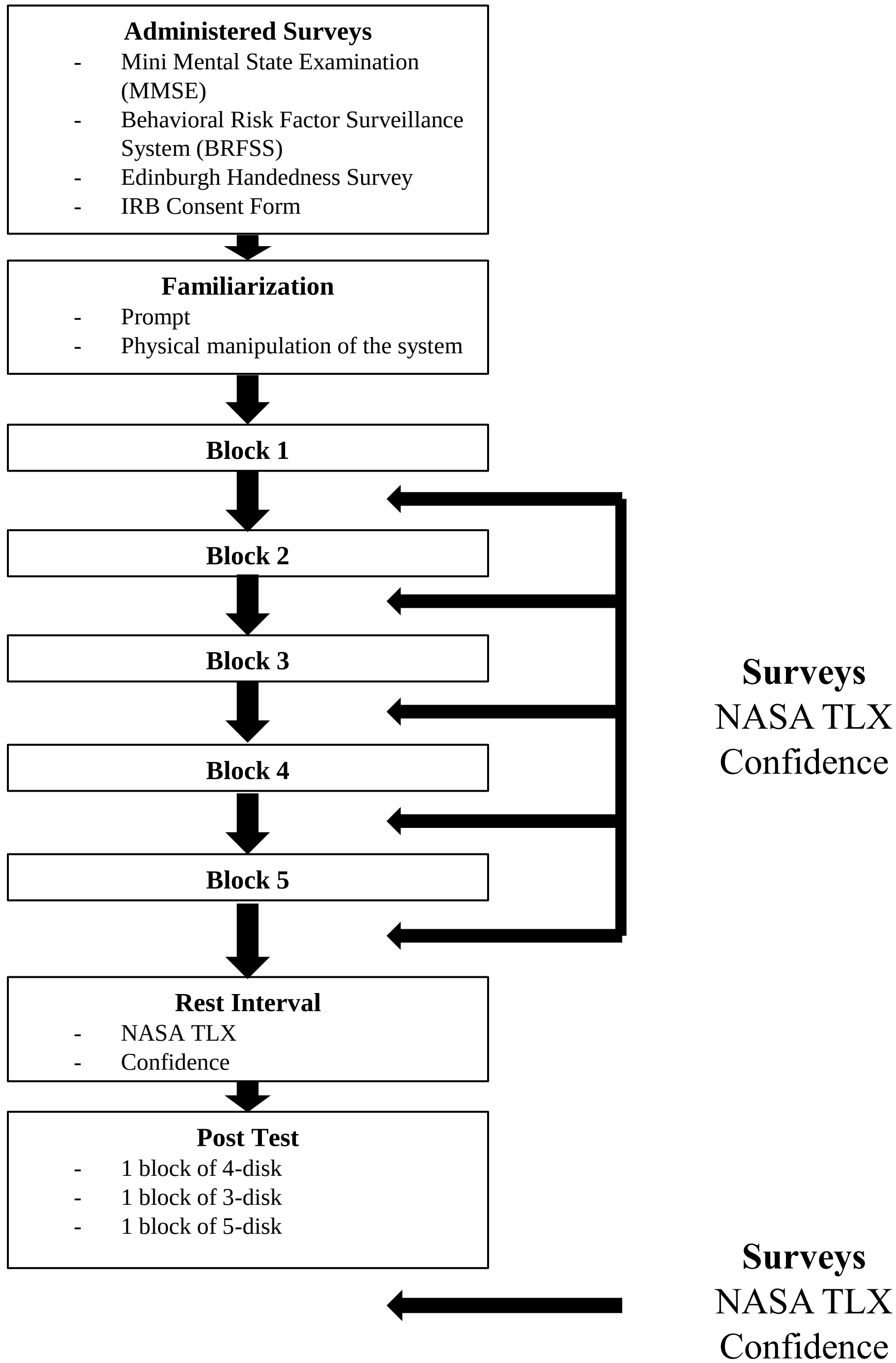

Figure 11. An illustration of the experimental procedure employing the Tower of Hanoi task. The practice phase is broken into five blocks, and after each block surveys to assess mental workload and confidence were administered.

III.c. Data Analysis

III.c.1 Action sequence modelling

Each action sequence of every participant was recorded via digital camera and then coded into the various actions that composed those sequences. Data analysis of the sequences was conducted using the novel computational approach presented in the first part of this work (see study 1 presented earlier; Hauge et al., *submitted to IJSR*). This computational method automates the calculation of LD for each complex motor sequence and produces graphical representations corresponding to the relative differences and similarities between the sequences.

Subaction Code	Natural Language Translation
'1a'	Ring 1 to peg A
'1b'	Ring 1 to peg B
'1c'	Ring 1 to peg C
'2a'	Ring 2 to peg A
'2b'	Ring 2 to peg B
'2c'	Ring 2 to peg C
'3a'	Ring 3 to peg A
'3b'	Ring 3 to peg B
'3c'	Ring 3 to peg C
'4a'	Ring 4 to peg A
'4b'	Ring 4 to peg B
'4c'	Ring 4 to peg C
'5a'	Ring 5 to peg A
'5b'	Ring 5 to peg B
'5c'	Ring 5 to peg C

Table 4. Table of complete motor alphabet for this task, *A**. Contains all possible moves for Tower of Hanoi task, including posttest conditions with a 5-ring tower.

Status Code	Natural Language Translation
‘pass’	Successful trial, within time limit and no rule violations
‘timeout’	Failed trial due to time limit lapse
‘rulebreak’	Failed trial due to rule violation

Table 5. Table of success/failure statuses used in MATLAB computational tool. These codes determine the initial categorization of the participant trials to be used in the program for generation of the various graphical analyses comparing subject performance.

Prior to data analysis, success and failure criteria were established to inform the LD calculations of difference between different trials. Successful trials must i) end in the goal state, ii) be completed within the time limit established in the study for each trial, and iii) follow the three rules of the Tower of Hanoi task. Each trial was coded as successful/failure using criteria of a new identifier within the MATLAB program so that the tool could seamlessly categorize trials based on their status and then by the subtype of this identifier (see table 5 above). The sequences were processed by this computational algorithm for computing the LD, the occurrence of its operators (insertion, deletion and substitution).

III.c.1.i. Statistical analysis

To assess the changes in performance throughout practice the LD and its three operators (insertion, deletion and substitution) as well as the MT were subjected to a one-way repeated measures ANOVA with the repetition on the factor BLOCK. In addition, to assess the modulations in mental workload and confidence during practice, each dimension of the NASA TLX and confidence surveys were subjected to the same one-way repeated measures ANOVA analysis. The Greenhouse-Geisser correction was employed if the assumption of sphericity was violated in each of the repeated measures ANOVAs. When appropriate, post-hoc analysis was conducted by employing Tukey's HSD. Also, in order to compare the LD for the five practice blocks to that for the optimal sequence (i.e., LD = 0) *t-tests* or *Wilcoxon signed rank* tests were employed depending if the assumption of normality was met or not. The significance criterion $\alpha = 0.05$ was employed for all tests. To assess the changes in performance while performing the ToH task under three levels of task difficulty during the post-test (i.e., transfer), the LD and its three operators (insertion, deletion and substitution) as well as the MT were subjected to a one-way repeated measures ANOVA with the repetition on the factor DIFFICULTY. Similarly, to assess the changes in level of mental workload and confidence when performing the task under different level of difficulty each dimension of the NASA TLX and confidence surveys were subjected to same one-way repeated measures ANOVA analysis. The Greenhouse-Geisser correction was employed if the assumption of sphericity was violated in each of the repeated measures ANOVAs. When appropriate, the Tukey's HSD post-hoc were employed. The significance criterion $\alpha = 0.05$ was employed for all tests.

III.c.2. Results

III.c.2.i. Changes in cognitive-motor performance throughout practice.

Analysis of results indicate that across all trials of practice, 88.8% of the trials were considered successful based upon the study's predetermined criteria of success. These successful trials included a variety of motor sequences that in some instances deviated from the most optimal solution path. More specifically, 56.3% of the successful trials employed suboptimal solution paths, while the remaining 43.7% of successful trials employed the optimal solution path. Practice trials that resulted in failures made up 11.2% of the trials, with 92.9% of failed trials a result of the trial time lapsing and the remaining 7.1% of failed trials a result of actions which violated the rules of the task space.

Levenshtein distance and its operators.

There was a significant main effect of BLOCK on LD across practice ($F(1.940, 36.862) = 7.431$, $\eta_p^2 = 0.281$, $p = 0.002$). Post-hoc analysis indicated a significant decrease of LD during practice between block 1 relative to block 3 ($p = 0.033$), block 4 ($p = 0.001$) and 5 ($p < 0.001$), as well as a significant reduction from block 2 to block 5 ($p = 0.002$). The same analysis revealed a significant main effect of BLOCK on substitutions ($F(2.289, 43.488) = 7.741$, $\eta_p^2 = 0.289$, $p = 0.001$) and insertions ($F(2.119, 40.263) = 5.489$, $\eta_p^2 = 0.224$, $p = 0.007$) in the motor sequences performed across blocks of practice. Post-hoc analysis revealed a significant reduction of substitutions from practice block 1 to blocks 3 ($p = 0.005$), block 4 ($p < 0.001$), and block 5 ($p < 0.001$), as well as a significant decrease between block 2 and block 5 ($p = 0.006$) of practice. There was also a significant reduction of the number of insertions from practice block 1 compared to blocks 4 ($p = 0.014$) and block 5 ($p < 0.001$), as well as a significant decrease between block 2 and block 5 ($p = 0.007$). In addition, the LD was significantly different from zero for all practice blocks, suggesting that the optimal sequence was not reached by the end of practice ($p<0.001$ for all comparisons). The changes in LD during practice were mainly due to modification of the number of substitution and insertion and changes in the number of deletions (Figure 12).

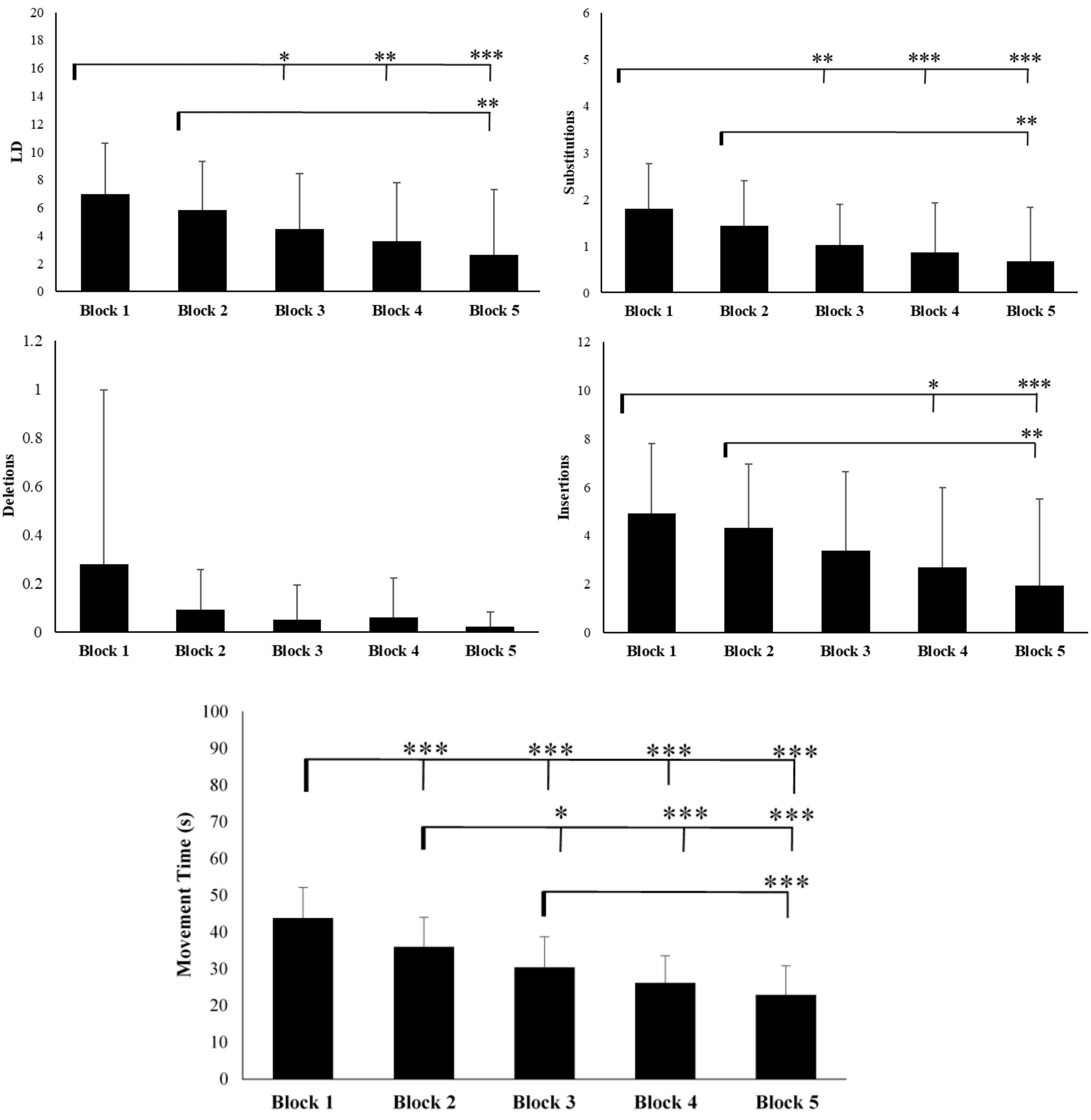

Figure 12. LD and its operators and MT across all five blocks of practice. From left to right, top to bottom: LD, Substitutions (S), Deletions (D), Insertions (I), Movement Time (MT). Significant p-values: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

Movement time.

There was a main effect of BLOCK for MT ($F(2.528, 48.038) = 38.398$, $\eta_p^2 = 0.669$, $p < 0.001$). Post hoc analyses indicated that compared to block 1 the MT was significantly reduced in blocks 2 ($p < 0.001$), block 3 ($p < 0.001$), block 4 ($p < 0.001$), and block 5 ($p < 0.001$). A significant decrease from practice block 2 to blocks 3 ($p = 0.022$), block 4 ($p < 0.001$) and block 5 ($p < 0.001$), as well as from block 3 to block 5 ($p < 0.001$) was observed.

NASA-TLX.

There was a main effect of BLOCK for the mental demand dimension throughout practice ($F(2.040, 38.754) = 6.839$, $\eta_p^2 = 0.265$, $p = 0.003$). Post hoc analyses indicated a significant reduction between block 1 and blocks 4 ($p = 0.002$) and block 5 ($p < 0.001$) as well as between blocks 2 and 5 ($p = 0.005$). Furthermore, a main effect of BLOCK was identified for the temporal demand ($F(2.314, 43.961) = 7.579$, $\eta_p^2 = 0.285$, $p = 0.001$), performance ($F(2.597, 56.191) = 10.567$, $\eta_p^2 = 0.357$, $p < 0.001$), effort ($F(2.052, 38.990) = 7.694$, $\eta_p^2 = 0.288$, $p = 0.001$), and frustration ($F(4,76) = 6.839$, $\eta_p^2 = 0.265$, $p = 0.003$) scale. There was no main effect of BLOCK for the physical demand dimension ($F(1.882, 34.616) = 1.067$, $\eta_p^2 = 0.053$, $p = 0.350$).

Post hoc analyses for each dimension indicated a significant reduction of the temporal demand (block 1 vs blocks 4: $p = 0.001$; block 1 vs. block 5: $p < 0.001$), effort (block 1 vs blocks 4: $p < 0.001$; block 1 vs. block 5: $p < 0.001$), and frustration (block 1 vs blocks 4: $p < 0.001$; block 1 vs. block 5: $p < 0.001$) dimension during practice. Also, a significant decrease between blocks 2 and block 5 in the temporal demand ($p = 0.003$), effort ($p < 0.001$) and frustration ($p = 0.013$) dimension was revealed. Finally, the effort ($p = 0.022$) and frustration ($p = 0.006$) subscales were significantly reduced in block 5 relative to block. Furthermore, post hoc analyses revealed an improved sense of performance between block 1 and block 4 ($p = 0.002$); block 1 and 5 ($p < 0.001$) as well as between blocks 2 and block 5 ($p = 0.003$), and between block 3 and block 5 ($p = 0.045$) (Figure 13).

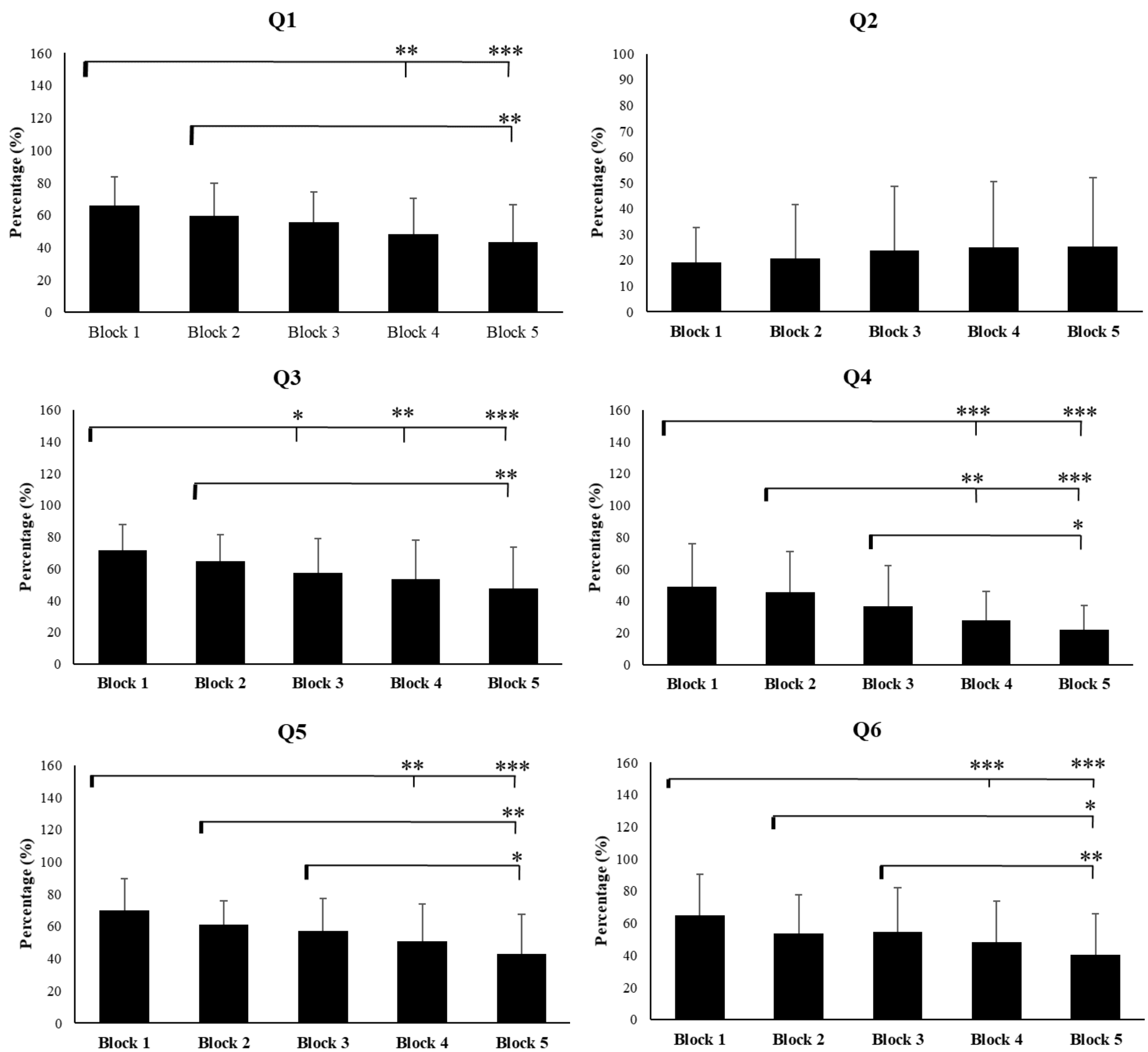

Figure 13. NASA TLX questionnaire results for all participants across blocks of practice. Left to right, top to bottom: Mental demand (Q1), Physical demand (Q2), Temporal demand (Q3), Performance (Q4), Effort (Q5), Frustration (Q6). Significance = *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

Confidence.

A main effect of BLOCK for participants' confidence to have performed the task well (Q4: $F(2.735, 51.968) = 19.063$, $\eta_p^2 = 0.501$, $p < 0.001$; Q4: $F(2.129, 40.453) = 14.577$, $\eta_p^2 = 0.434$, $p < 0.001$). Post hoc analyses indicate a significant increase in confidence to perform efficiently (Q3) the task i) in blocks 3, 4 and 5 compared to block 1 ($p < 0.001$ for all comparisons), ii) in block 4 ($p = 0.026$) and block 5 ($p < 0.001$) relative to block 1; iii) in block 5 compared to block 3 ($p = 0.006$). Finally, there was a significant increase in confidence to perform the task well (Q4) i) in blocks 3, 4 and 5 relative to block 1 ($p < 0.003$ for all comparisons); ii) in blocks 4 ($p = 0.00312$) and 5 ($p < 0.001$) compared to block 2 and iii) in block 5 relative to block 3 ($p < 0.01$) of practice.

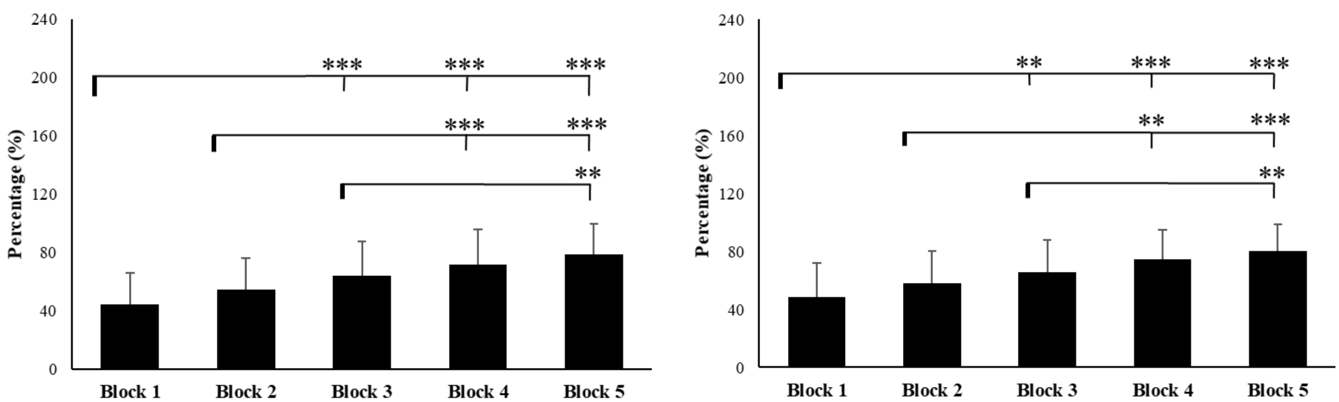

Figure 14. Assessment on how individuals confident that they performed the task as accurately and efficiently as possible (left panel; Q1) and as well as possible (right panel; Q2) for each practice blocks. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

Primacy/recency.

The exploratory analysis of the primacy and recency effect across blocks of practice revealed an overall trend of primacy and recency occurring throughout practice (see Figure 15). Primacy-recency effects tended to be most prominent during early relative to middle practice. By late practice a minimal number of LD operators is reached and all of which occur during the 5th step of the reference motor sequence.

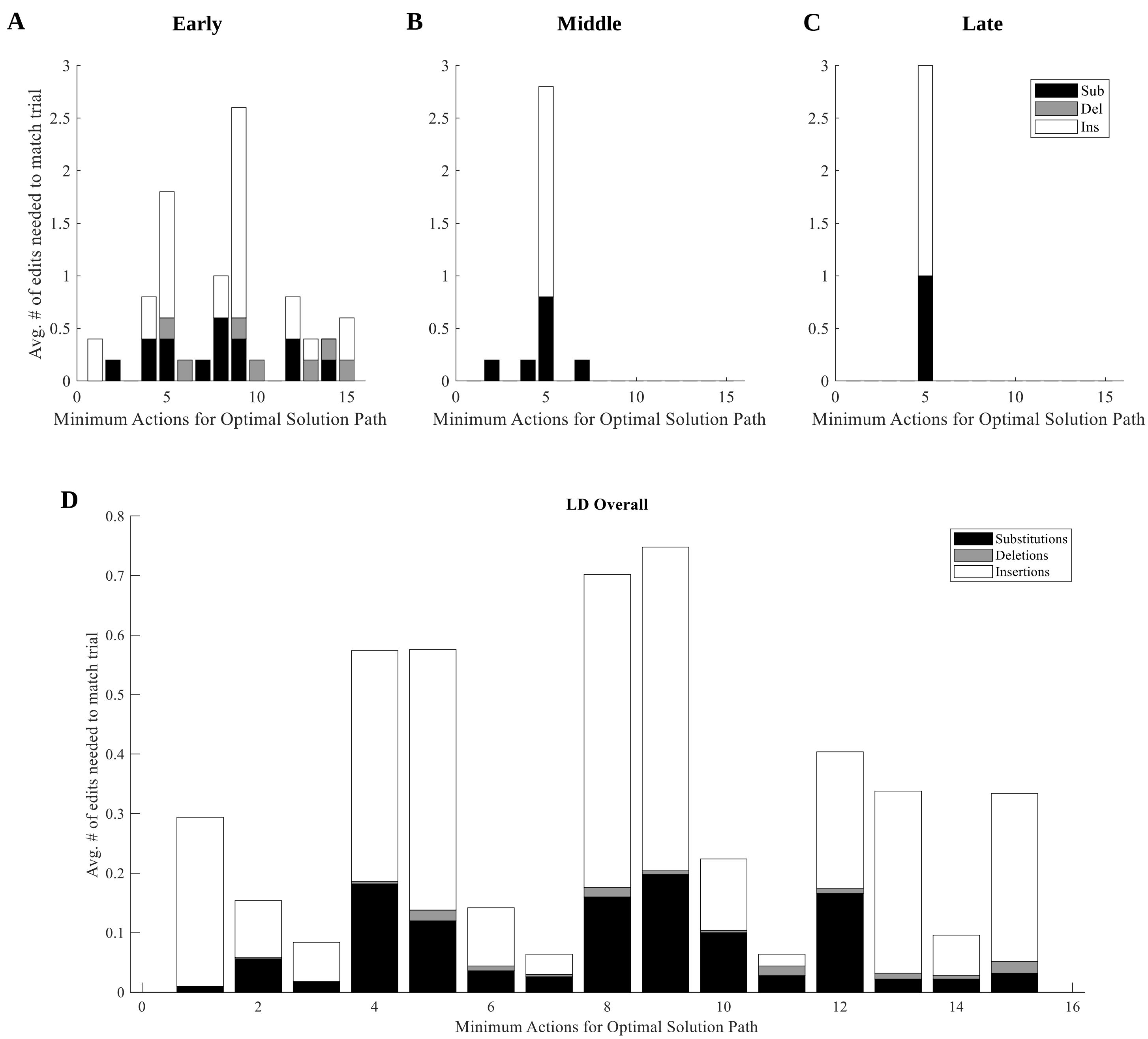

Figure 15. Primacy-recency of LD in early block of practice (A), middle (B), and late (C) blocks of practice. Primacy-recency of Levenshtein edit distance and the three operators across all blocks of practice (D).

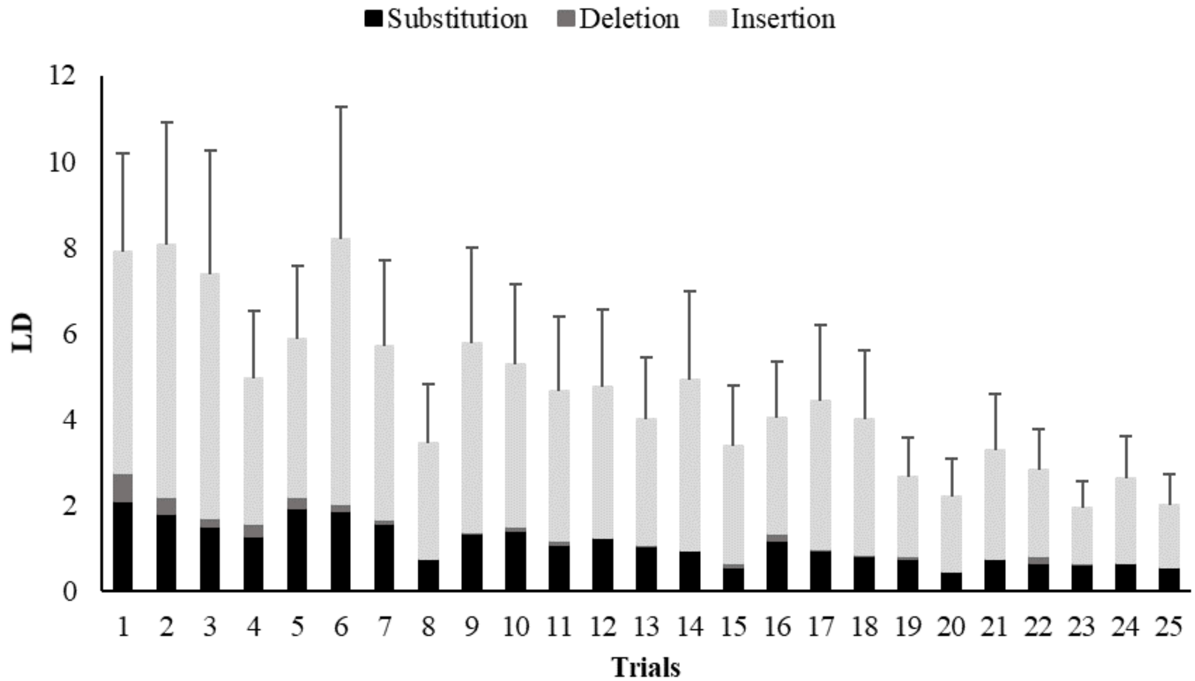

Figure 16. Average LD across all subjects throughout each trial of practice.

III.c.2.ii. Post-Test.

Analysis of performance during each post-test condition indicates that subject performance in the 3-ring and 4-ring conditions were overwhelmingly successful (3-ring: 98% successful; 4-ring: 94% successful), and subject performance in the 5-ring post-test condition was also successful for a majority of the total trials (5-ring: 65% successful). Post-testing did not guarantee a reduction in LD as a result of the 4-ring practice session. In the 3-ring post-test condition, 89.8% of trials employed the optimal solution; 10.2% of the remaining successful trials used a suboptimal solution path. Interestingly, only 1 of the successful trials (1.1%) employed the optimal solution path in the 4-ring post-test condition compared to the other successful trials (98.9%). Comparatively, 9.2% of the successful trials in the 5-ring post-test condition used the optimal solution, while the remaining 90.8% of the successful trials used a suboptimal solution path.

Failed trials in each of the post-test conditions varied in their variety of LD operators present as well as the criteria which invalidated the trial. The 3-ring condition contained only 2 failed trials, both of which occurred because the trial time elapsed (average: LD= 12; S= 2.5; D= 0.5; I = 9). The 4-ring post-test condition contained 6 failed trials: 3 trials contained rule violations (average: LD= 15; S= 5; D= 0; I= 10), and the other 3 trials timed out (average: LD=13.67; S=3.33; D=0.333; I=10). Last, the 5-ring post-test condition contained 35 total failed trails: 7 trials contained rule violations (average: LD= 23; S= 4.14; D= 1.57; I= 17.29), while the remaining 28 trials timed out (average: LD= 30.32; S= 5.79; D= 1.25; I= 23.29).

Levenshtein distance and its operators.

There was a main effect of DIFFICULTY for the LD ($F(1.159, 22.030) = 88.410$, $\eta_p^2 = 0.823$, $p < 0.001$). Post hoc analyses indicated a significant increase for the practiced and hard relative to the easy condition ($p < 0.001$, for all comparisons). There was also a main effect of DIFFICULTY for the substitutions ($F(2,38) = 57.084$, $\eta_p^2 = 0.750$, $p < 0.001$). Post hoc analyses indicated a significant increase of the number of substitutions with the practiced and hard compared to the easy condition ($p < 0.001$), whereas a significant decrease was observed for the hard compared the practiced condition ($p = 0.007$). A tendency was observed for DIFFICULTY for the number of deletions ($F(1.033, 19.630) = 4.024$, $\eta_p^2 = 0.175$, $p = 0.058$). Post hoc analyses indicated that the number of deletion tended to increase in the practiced and hard compared to the easy condition ($p < 0.001$). There was a main effect of DIFFICULTY for the number of insertions ($F(1.358, 25.804) = 71.171$, $\eta_p^2 = 0.789$, $p < 0.001$). Post hoc analyses indicated a significant increase of the number of insertion in the practiced ($p = 0.018$) and hard ($p = 0.049$) relative to the easy condition.

Movement time.

There was a main effect of DIFFICULTY for the MT ($F(1.373, 26.088) = 265.967$, $\eta_p2 = 0.933$, $p < 0.001$). Post hoc analyses indicated a significant increase of MT from the easy to practiced and to hard condition ($p < 0.001$, for all comparisons).

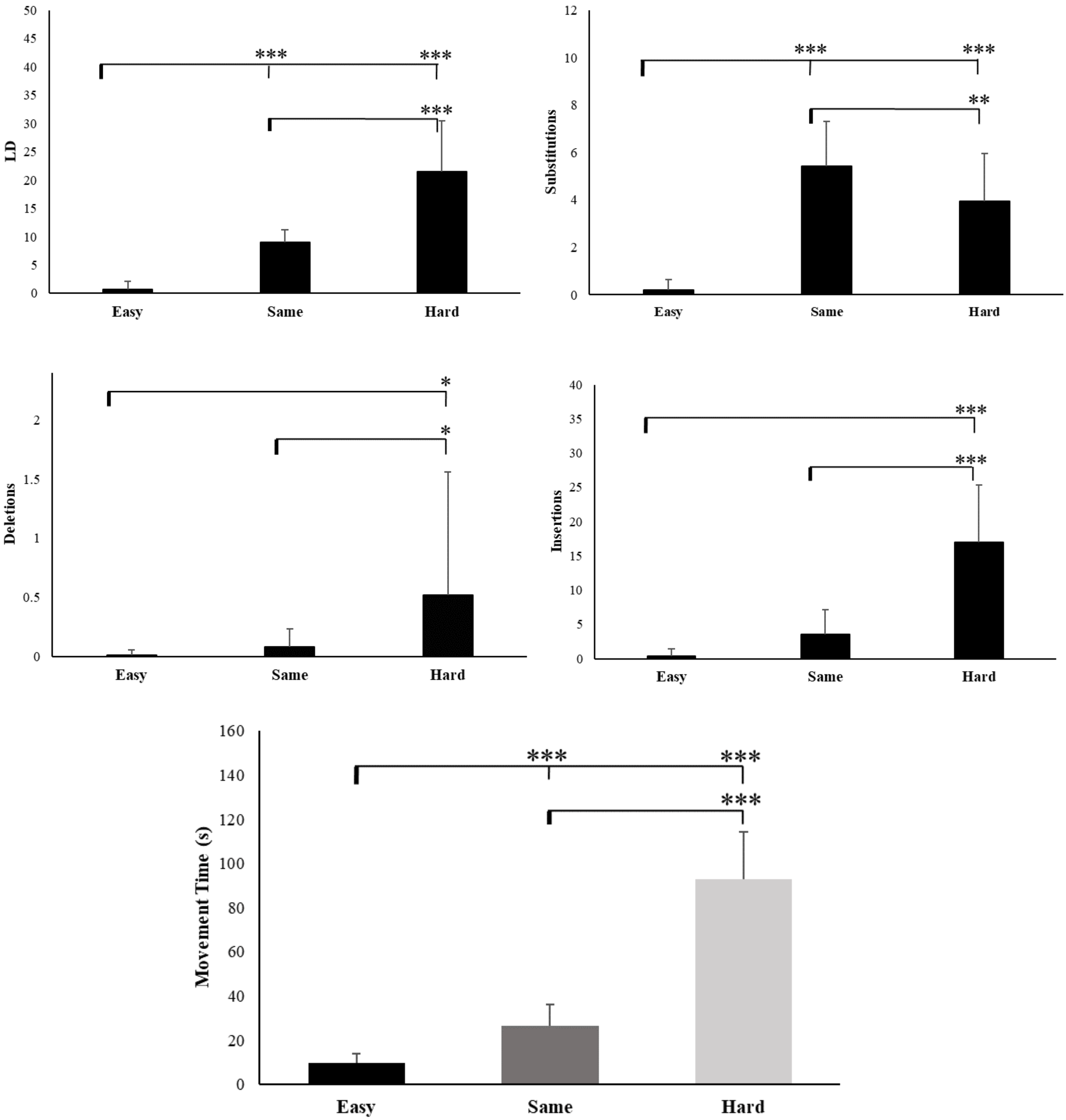

Figure 17. Levenshtein distance operators and movement time for each post-test condition. The easy, practiced and hard condition required to solve the ToH task with 3, 4 and 5 disks, respectively. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

NASA-TLX.

There was a main effect of DIFFICULTY for the mental ($F(2, 38) = 43.176$, $\eta_p^2 = 0.694$, $p < 0.001$), physical ($F(2, 38) = 4.503$, $\eta_p^2 = 0.192$, $p = 0.018$) and temporal ($F(2, 38) = 18.048$, $\eta_p^2 = 0.487$, $p < 0.001$) demand as well as performance ($F(2, 38) =21.207$, $\eta_p^2 = 0.527$, $p < 0.001$), effort ($F(2, 38) = 30.232$, $\eta_p^2 = 0.614$, $p < 0.001$) and frustration ($F(2, 38) = 17.008$, $\eta_p^2 = 0.472$, $p < 0.001$) dimensions of the NASA TLX survey. Post hoc analyses indicated a significant increase of mental ($p < 0.001$), physical ($p = 0.006$) and temporal ($p < 0.001$) demands as well as performance ($p < 0.001$), effort ($p < 0.001$), and frustration ($p < 0.001$) scales. There was also a significant elevation of mental demand ($p = 0.002$) and effort ($p = 0.005$) for the practiced compared to the easy condition. There was also a significant elevation of mental, temporal, effort and frustration as well as a reduction of perceived performance in the hard compared to the practiced condition ($p < 0.001$, for all comparisons).

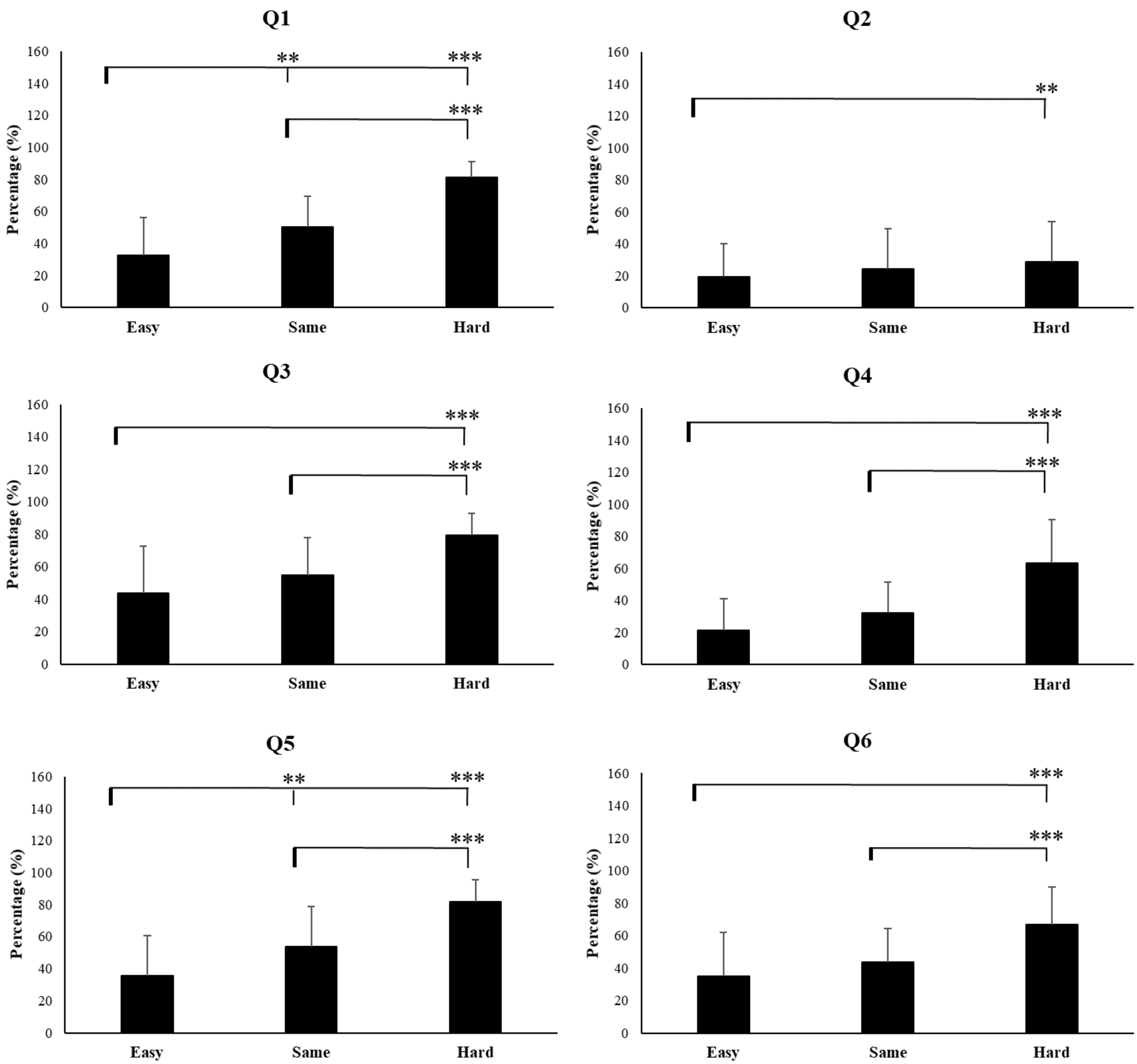

Figure 18. NASA-TLX survey results for each post-test condition. Left to right, top to bottom: Mental demand (Q1), Physical demand (Q2), Temporal demand (Q3), Performance (Q4), Effort (Q5), Frustration (Q6). *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

Confidence.

There was a main effect of DIFFICULTY for the confidence in performing the task efficiently (Q1: $F(2,38) = 21.184$, $\eta_p^2 = 0.527$, $p < 0.001$; Q2: $F(2,38) = 25.315$, $\eta_p^2 = 0.571$, $p < 0.001$). Post hoc analyses indicate a significant reduction in confidence to perform well and efficiently the task in hard compared to the practiced and easy condition ($p < 0.001$, all comparisons considered for Q1 and Q2.

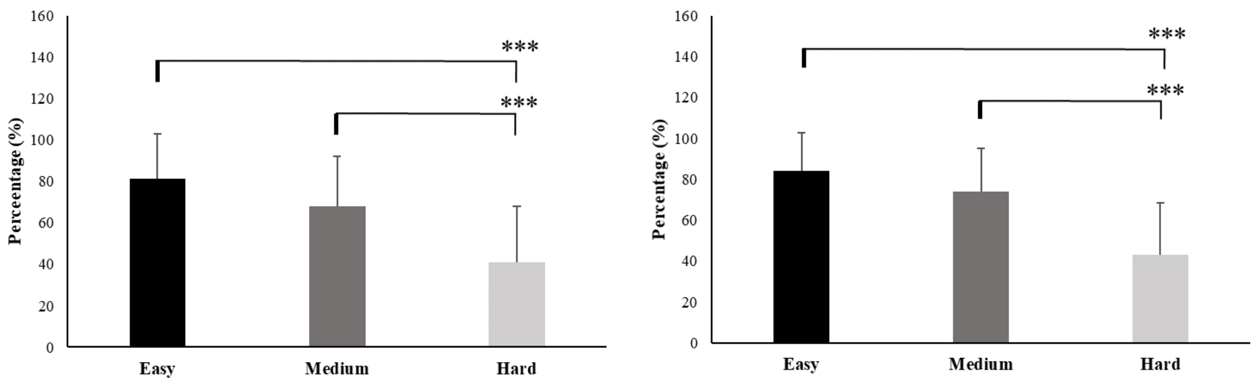

Figure 19. Assessment on how individuals are confident that they have performed the task as accurately and efficiently as possible (left panel; Q1) and as well as possible (right panel; Q2) under the easy, practiced and hard level of challenge. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

IV. Discussion

IV.a. Performance dynamics

IV.a.1. Motor performance, mental workload and confidence

As expected, as individuals practiced the action sequence: i) the motor performance was improved as indicated by a reduction of LD and of the number of insertion and substitutions (but not deletion) as well as MT; ii) an attenuation of the level of mental workload and iii) an elevation of the level of individual's confidence in performing the task well. This confirms and extend previous effort to complex action sequence suggesting that as individuals become more proficient at performing the action sequence task, the performance becoming more automatic the mental demand is progressively reduced and the confidence is enhanced (Gentili et al. 2015, 2013; Rietschel et al. 2014; Ong et al. 2015, 2017; Trempe et al. 2012; Stevens et al 2012). However, closer inspection of the results suggest that the velocity at which these changes occur is different for the structure of the high-level plan, the time to perform the sequence, the mental demand, and the confidence.

First, although the level of mental demand decreased throughout practice, this reduction occurred at a slower rate than that observed for the change of the structure of the high-level motor plan. This was indicated by the fact that while a significant reduction LD as well as the number of insertion and substitution was reached as early as block 3, a significant reduction of mental demand only occurred at the two last following blocks (blocks 4 and 5). This suggests that in block 3 while a better high-level motor plan was generated that was closer to the optimal solution (i.e., the one with a minimum of actions), a substantial amount of mental resources was still engaged. Such a difference in dynamics confirms and extends previous work that has examined motion planning during center-out reaching movements where the behavior improved faster than the refinement of brain dynamics (Gentili et al., 2011, 2015). It must be noted that a similar lag was observed between the changes in mental demand and MT which was significantly improved as soon as the second practice block. In addition, although some additional assessment should be conducted it must be noted that the changes in LD (and its insertion and substitution operator) and those in MT revealed different patterns of improvement. Namely, MT starts to decrease as early as in the two first blocks and is still being reduced from middle to late practice (i.e., block 3 vs. block 5). However, the LD along with the number of insertions and substitutions is reduced (i.e., shows a convergence toward the optimal sequence) only during comparison of late and early practice blocks. This may suggest that at an early practice stage the execution time of the action sequence is reduced likely since individuals moves faster the disks without however necessarily improving their high-level motor plan in order to converge towards the optimal sequence.

Similarly, the fact that the MT still decreases from middle to late practice whereas this is not the case neither for the LD nor for the number of insertions and substitutions may suggest that the high-level motor plan identified by the individuals during middle practice (block 3) is considered acceptable, and from then the overall performance improvement would be based on moving the disk faster with this specific motor plan. Another possibility would be that more trials are needed to further improve this high-level motor plan. Finally, although the mental demand dimension of the NASA TLX was of primary interest since the most related to mental effort (Bittner et al., 1989; Hendy et al., 1993; Hart, 2006; Van Gog and Paas, 2008; Kujala, 2012; Ayaz et al., 2013; Akizuki and Ohashi, 2015; Shaw et al. 2018) this result is also consistent with performance, effort and frustration dimensions of the NASA TLX. The temporal dimension is the only metric which already shows substantial improvement at block 3, although this is likely influenced with the rapid changes in MT. The pattern of changes in motor performance (reduction of LD, number of insertions/substitutions and MT) and confidence were more similar as suggested by the modulation from middle to early practice stage. As such, both motor performance and confidence improved faster compared to the mental demand. Thus, it is possible that, upon reaching a consistent solution that the individuals are more confidence in their capability to execute efficiently the task by the third practice block which may promote performance improvement and in turn facilitate a reduction of mental demand in the next (fourth) practice block.

IV.a.2 Levenshtein distance and its operators

Contrary to the original hypothesis surrounding motor performance, subjects did not converge on the optimal action sequence (LD = 0) by the last practice block. Of the 441 successful trials of practice, more than half of the trials (56.01%) employed a suboptimal motor sequence. Thus, the fact that average LD for the last practice block was significantly different from zero combined with a consistent reduction in MT may suggest that individuals converged first to a local minimum of LD while still on reducing the time to completion for each trial. This may be explained by a practice strategy with two processes having different dynamics where the time requirements to complete the task. A low-level sensorimotor process leads to a fast decrease of movement time by moving the disks rapidly which is complemented by a slower process employed to build the structure of the high-level plan to perform the action sequence with a minimum of move. However, additional analyses should be conducted to further examine this hypothesis.

Generally, the LD for successful trials was lower than rulebreak and timeout trials, consistent with the hypothesis that LD can be related to overall performance. In addition, the composition of LD in successful trials and both types of failed trials (‘rulebreak’ and ‘timeout’) differs proportionately. Namely, for successful trial, the LD_{Suc} was mainly driven by the number of substitution (25.63%) and insertion (73.7%) and almost not to deletion (0.66%). In addition, the composition of the LD for the timeout and successful trials were similar although an almost 6-fold increase of deletion (S = 23.26%; D = 3.78%; I = 72.96%). Importantly, compared to the two other conditions, the rulebreak condition revealed a composition of the LD having an increased proportion of deletions (28.125%) and decreased proportions of insertions (53.12%) and substitution (18.75%). This finding is consistent with previous work which suggested that deletion was a marker of performance failure in an action sequence (Hauge et al. *submitted*). It must be noted that the similar composition of the LD for the successful and timeout trials suggests that for these trials individual may have been able to successfully perform the task if no time limit was imposed. If that assumption is correct, then it is possible that a better assessment would be to consider the ‘rulebreak’ trials as the primary criteria of failure in task performance.

Primacy and recency effects were present during practice of the complex motor sequence, confirming previous work surrounding completion of motor tasks and memory (Deese & Kaufman, 1957; Murdoch, 1962). What is interesting to note is that the overall LD occurring throughout the actions of the optimal solution path during practice show elements of the sequence with a very high LD followed immediately by elements of the sequence with a much lower LD (see Figure 15D). It is possible that the steps in which there is a higher LD are critical movements in the sequence which are frequently missed or reorganized and thus lead to deviation from the optimal solution path. By the late block of practice, there were no deletions present, and only one step of the optimal solution path revealed the presence of insertions and substitutions.

IV.b. Post-test

IV.b.1. Motor performance, mental workload and confidence

When comparing all three levels of challenge, as difficulty increased, the LD, MT, mental demand and sense of effort increased from the easy, practiced and hard condition. This suggests that as the task difficulty increased, the structure of the action sequence produced was further away from the optimal solution while its execution took more time and more mental effort. It must be noted that the same pattern was observed for the deletion, insertion, temporal demand, frustration and confidence, but this difference was only observed between the two extreme (easy and hard) conditions of transfer. This may suggest that the LD, MT and mental demand would be more sensitive to a gradual elevation of the level of task difficulty compared to the insertion and deletion operators. However, the transfer to the easy condition did not necessarily reduce the number of insertions and deletions, the sense of frustration, and enhance the confidence of being able to perform the task well. This pattern is consistent with that observed for the success rate for the easy (98%), practiced (94%) and hard condition (65%).

IV.b.2 Levenshtein distance and its operators

Although no clear pattern appeared between the conditions of transfer, the successful action sequence had a LD mainly driven by substitution and insertion and no deletion (S = 41%; D = 0; I = 58.97%). However, the proportion of deletion increased for failed trials in particular for the hard condition. Namely, while the LD composition in this condition for the successful (S= 17.76%; D= 0.53%; I= 81.58%), timeout (S= 19.08%; D= 4.12%; I= 76.8%) and rulebreak (S= 18.01%; D= 6.83%; I= 75.16%), trials was relatively similar he deletion operator tended to increase for the action sequence failure. This is in agreement with previous work that the deletion may be a performance marker more sensitive to action sequence failure (Hauge et al. *submitted*).

IV.c. Applications

IV.c.i Clinical settings

This present work has potential benefits to individuals within a clinical population having cognitive-motor deficiencies and in particular with high-level motor planning for performing complex action sequences. In particular, the use of LD and its operators can provide performance markers of complex motor sequences executed in the context of performance, practice or re-learning over multiple training sessions. For instance, the LD and its operators could be employed to track performance across rehabilitative sessions and also identify any part of the action sequence upon which the rehabilitation should focus. This work can also contribute to develop diagnostic tools for patients by providing additional metrics for performance beyond completion time of the task. The nuances of the motor behavior can be more thoroughly tracked using this protocol and may be used to enhance therapies for these populations (Sanford, Moreland, Swanson, Stratford, & Gowland, 1993; Bosecker, Dipietro, Volpe, & Igo Krebs, 2010).

IV.c.ii. Complex skill training

This work can also be applied to training and proficiency assessment of real-world skills which involve complex action sequences for humans in high risk performance activities. For instance, several studies have identified the importance of monitoring the mental demand during activities of varying levels of difficulty, most notably in United States Armed Forces pilots and soldiers whose training and skill assessment are critical to occupational safety and efficiency

(Jacquess et al. 2017; Gentili et al. 2018; Berka et al. 2007; Hankins & Wilson, 1998; Kacer et al., 2018). In particular, the proposed computational approach can identify insertion, deletion or substitution of actions during performance as well as throughout practice, thus informing whether the motor sequences generated by individuals are optimal or suboptimal. This, combined with the assessment of mental workload, can inform the efficiency at which an action sequence is executed by individuals. As such, in turn this can inform training techniques that can promote the performance of optimal action sequence while minimizing mental workload and thus promoting performance efficiency (Vink et al., 1995; Recarte & Nunes, 2003).

IV.c.iii. Human-robot interaction

The flexibility and adaptability exhibited by human performance during practice can inform human-robot interaction studies. While sensorimotor systems have reached a certain level of sophistication in robotic systems, their lack of cognitive functioning does not allow them to generate high-level motor plans for the performance ability and flexibility necessary to navigate their environments freely and smoothly. Bridging the gap between repeatable and consistent performance and adaptation is a nuanced task, and it is possible that this work can contribute to provide the approach necessary to identify faulty actions and extraneous missteps that may compromise task performance. Evaluating human performance during high level complex motor tasks followed by subsequent coding of performance through this LD algorithm could inform robotic interactions with humans that would maximize performance outcomes and comfort within the environment for the human user (Scholtz, 2003; Scholtz, Young, Drury, & Yanco, 2004).

V.d. Limitations

The first limitation of note was the use of surveys to evaluate the changes of the level of mental workload during practice. Besides being subjective by design, the surveys could not be used on a single trial basis and instead were used with five 5-trial block periods which provide a limited refinement in the analysis of these mental workload changes. Future work could address this limitation by employing Electroencephalography (EEG) to assess the cortical dynamics. A second limitation may be the use of a 60 second time limit for each trial. Although a time limit for each trial was assumed to promote task engagement and speed to complete the motor sequence, the similarity in the composition of the LD in the successful and failed trials resulting from the time elapsing suggests that such a constraint could be modified either by removing this 60 second time limit or by extending this time limit for each trial completion which may isolate the failed trials due solely to the violation of task rules and thus provide a slightly more refined view of the results. Another limitation of this study is the sample size of the group of participant in this study and thus these results have to be considered in light of this limitation.

VI. General conclusion

This work aimed to examine the concurrent dynamics of motor performance, mental workload and confidence and throughout practice of a complex motor sequence. To reach this aim, first a computational approach to provide metrics which had to be concise (e.g., distance, scalar) while still providing detailed information of the performed action sequence. Thus, the first study reported here proposed a LD-based computational approach able to assess the performance of action sequences in a versatile context where the task goal could be reached in a flexible manner by a human or robotic performer. The results revealed that this novel approach can provide different yet complementary metrics to those already existing to assess complex task performance by quantifying the differences between action sequences executed under various conditions (e.g., outcome; performed by humans or humanoid robots). Then, in a second study this novel computational approach was deployed to examine the concurrent changes in cognitive-motor performance along with the modulations of mental workload and confidence throughout practice of a complex motor sequence to solve the ToH task. The results revealed that during practice, the motor performance was enhanced (decrease of LD, insertion and substitution as well as movement time) whereas the level of mental workload and confidence decreases and increases, respectively. It must be noted that by late practice the high-level plans generated to perform the corresponding action sequences were not optimal although being executed more rapidly. This possibly suggests a different dynamic between the low-level sensorimotor control, which rapidly improves movement duration, and the formation of high-level motor plans. Future work could extend this current effort by considering concurrent brain dynamics using EEG and also include a longer practice period extended post-testing to assess long-term skill acquisition. Overall, this work complements prior work that have examined complex action sequence executed in humans and humanoid robots during cognitive-motor performance and practice and can potentially inform not only human cognitive-motor mechanisms in healthy and clinical population but also human-robots interactions.

VII. Acknowledgements

The first study of this thesis project was funded by The Office of Naval Research (N000141310597).

VIII. References

Akizuki, K., & Ohashi, Y. (2015). Measurement of functional task difficulty during motor learning: What level of difficulty corresponds to the optimal challenge point? *Human Movement Science*, *43,* 107-117. http://dx.doi.org/10.3389/fnhum.2013.00871

Ansuini, C., Santello, M., Massaccesi, S., & Castiello, U. (2006). Effects of end-goal on hand shaping. *Journal of neurophysiology*, *95*(4), 2456-2465.

Ashe, J. (1997). Erratum to "Force and the motor cortex": [Behavioural Brain Research 86 (1997) 1–15] 1PII of original article: S0166-4328 (96) 00145-31. *Behavioural brain research*, *87*(2), 255-269.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child neuropsychology*, *8*(2), 71-82.

Anderson, J. R., Albert, M. V., & Fincham, J. M. (2005). Tracing problem solving in real time: fMRI analysis of the subject-paced Tower of Hanoi. *Journal of cognitive neuroscience*, *17*(8), 1261-1274.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122-147.

Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., & Craven, P.L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, *78*(5), B231-244.

Borys, S. V., Spitz, H. H., & Dorans, B. A. (1982). Tower of Hanoi performance of retarded young adults and nonretarded children as a function of solution length and goal state. *Journal of Experimental Child Psychology*, *33*(1), 87-110.

Bosecker, C., Dipietro, L., Volpe, B., & Igo Krebs, H. (2010). Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke. *Neurorehabilitation and neural repair*, *24*(1), 62-69.

Bosse, H. M., Mohr, J., Buss, B., Krautter, M., Weyrich, P., Herzog, W., . . . Nikendei, C. (2015). The benefit of repetitive skills training and frequency of expert feedback in the early acquisition of procedural skills. *BMC Medical Education*, *15*(22), 1-10. http://dx.doi.org/10.1186/s12909-015-0286-5

Bryson, A. E. (1996). Optimal control-1950 to 1985. *IEEE Control Systems*, *16*(3), 26-33.

Bullock, D., & Grossberg, S. (1988). Neural dynamics of planned arm movements: emergent invariants and speed-accuracy properties during trajectory formation. Psychological review, 95(1), 49.

Bullock, D., Grossberg, S., & Guenther, F. H. (1993). A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm. *Journal of Cognitive Neuroscience*, *5*(4), 408-435.

Bullock, D., Bongers, R. M., Lankhorst, M., & Beek, P. J. (1999). A vector-integration-to-endpoint model for performance of viapoint movements. *Neural Networks*, *12*(1), 1-29.

Butchibabu, A., Sparano-Huiban, C., Sonenberg, L., & Shah, J. (2016). Implicit coordination strategies for effective team communication. *Human factors*, *58*(4), 595-610.

Casartelli, L., Federici, A., Cesareo, A., Biffi, E., Valtorta, G., Molteni, M., ... & Borgatti, R. (2017). Role of the cerebellum in high stages of motor planning hierarchy. *Journal of neurophysiology*, *117*(4), 1474-1482.

Chan, R. C. K., Chen, E. Y. H., Cheung, E. F. C., Chen, R. Y. L., & Cheung, H. K. (2004). Problem–solving ability in chronic schizophrenia. *European archives of psychiatry and clinical neuroscience*, *254*(4), 236-241.

Chan, R. C., Wang, Y. N., Cao, X. Y., & Chen, E. Y. (2010). Contribution of working memory components to the performance of the Tower of Hanoi in schizophrenia. *East Asian Archives of Psychiatry*, *20*(2), 69.

Chen, Y. P., Keen, R., Rosander, K., & Von Hofsten, C. (2010). Movement planning reflects skill level and age changes in toddlers. *Child development*, *81*(6), 1846-1858.

Chiviacowsky, S., Wulf, G., & Lewthwaite, R. (2012). Self-controlled learning: The importance of protecting perceptions of competence. *Frontiers in Psychology*, *3*(458), 1-8. http://dx.doi.org/10.3389/fpsyg.2012.00458

Chiviacowsky, S. (2014). Self-controlled practice: Autonomy protects perceptions of competence and enhances motor learning. *Psychology of Sport and Exercise*, *15*(5), 505-510. http://dx.doi.org/10.1016/j.psychsport.2014.05.003

Claxton, L. J., Keen, R., & McCarty, M. E. (2003). Evidence of motor planning in infant reaching behavior. *Psychological Science*, *14*(4), 354-356.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*, 715-730. http://dx.doi.org/10.1037/0022-0663.88.4.715.

Dechent, P., Merboldt, K. D., & Frahm, J. (2004). Is the human primary motor cortex involved in motor imagery?. *Cognitive Brain Research*, *19*(2), 138-144.

Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. Journal of Experimental Psychology, *54*, 180-187.

Du, Y., Valentini, N. C., Kim, M. J., Whitall, J., & Clark, J. E. (2017). Children and adults both learn motor sequences quickly, but do so differently. *Frontiers in psychology*, *8*.

Egan, D. E., & Greeno, J. G. (1974). Theory of rule induction: Knowledge acquired in concept learning, serial pattern learning, and problem solving.

Fabbri-Destro, M., Cattaneo, L., Boria, S., & Rizzolatti, G. (2009). Planning actions in autism. *Experimental Brain Research*, *192*(3), 521-525.

Fireman, G. (1996). Developing a plan for solving a problem: A representational shift. *Cognitive Development*, *11*(1), 107-122.

Flanagan, J. R., Bowman, M. C., & Johansson, R. S. (2006). Control strategies in object manipulation tasks. *Current opinion in neurobiology*, *16*(6), 650-659.

Flash, T., & Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. Journal of neuroscience, 5(7), 1688-1703.

Freedman, R. G., & Zilberstein, S. (2017). Integration of Planning with Recognition for Responsive Interaction Using Classical Planners. In *AAAI* (pp. 4581-4588).

Gentili, R., Cahouet, V., & Papaxanthis, C. (2007). Motor planning of arm movements is direction-dependent in the gravity field. *Neuroscience*, *145*(1), 20-32.

Gentili, R. J., Bradberry, T. J., Oh, H., Hatfield, B. D., & Contreras Vidal, J. L. (2011). Cerebral cortical dynamics during visuomotor transformation: Adaptation to a cognitive-motor executive challenge. *Psychophysiology*, *48*(6), 813-824.

Gentili, R. J., Oh, H., Shuggi, I. M., Goodman, R. N., Rietschel, J. C., Hatfield, B. D., & Reggia, J. A. (2013, July). Human-robotic collaborative intelligent control for reaching performance. In *International Conference on Augmented Cognition* (pp. 666-675). Springer, Berlin, Heidelberg.

Gentili, R. J., Rietschel, J. C., Jaquess, K. J., Lo, L.-C., Prevost, M., Miller, M. W., . . . Hatfield, B. D. (2014). Brain biomarkers based assessment of cognitive workload in pilots under various task demands. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 5860-5863. http://dx.doi.org/10.1109/EMBC.2014.6944961

Gentili, R. J., Oh, H., Huang, D. W., Katz, G. E., Miller, R. H., & Reggia, J. A. (2015a). A Neural architecture for performing actual and mentally simulated movements during self-intended and observed bimanual arm reaching movements. *International Journal of Social Robotics*, *7*(3), 371-392.

Gentili, R. J., Shuggi, I. M., King, K. M., Oh, H., & Shewokis, P. A. (2015b). Cognitive-motor processes during arm-reaching performance through a human body-machine interface. *Lecture Notes in Computer Science, 9183,* 381-392. http://dx.doi.org/10.1007/978/-3-319-20816-9_36. Part of chapter in Foundations of Augmented Cognition.

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 1416-1419.

Gerardin, E., Sirigu, A., Lehéricy, S., Poline, J. B., Gaymard, B., Marsault, C., ... & Le Bihan, D. (2000). Partially overlapping neural networks for real and imagined hand movements. *Cerebral cortex*, *10*(11), 1093-1104.

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in “planning” functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia*, *33*(5), 623-642.

Goldenberg, G., Wimmer, A., Auff, E., & Schnaberth, G. (1986). Impairment of motor planning in patients with Parkinson’s disease: evidence from ideomotor apraxia testing. *Journal of Neurology, Neurosurgery & Psychiatry*, *49*(11), 1266-1272.

Goldman-Rakic, P. S. (1987). Development of cortical circuitry and cognitive function. *Child development*, 601-622.

Gordon, J., Ghilardi, M. F., Cooper, S. E., & Ghez, C. (1994b). Accuracy of planar reaching movements. II. Systematic extent errors resulting from inertial anisotropy. *Experimental brain research*, *99*(1), 112-130.

Haggard, P. (1998). Planning of action sequences. *Acta Psychologica*, *99*(2), 201-215.

Hanakawa, T., Dimyan, M. A., & Hallett, M. (2008). Motor planning, imagery, and execution in the distributed motor network: a time-course study with functional MRI. *Cerebral cortex*, *18*(12), 2775-2788.

Hanakawa, T., Immisch, I., Toma, K., Dimyan, M. A., Van Gelderen, P., & Hallett, M. (2003). Functional properties of brain areas associated with motor execution and imagery. *Journal of neurophysiology*, *89*(2), 989-1002.

Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, space, and environmental medicine*, *69*(4), 360-367.

Harrison, J., Izzetoglu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., . . . Onaral, B. (2014). Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. *IEEE Transactions on Human-Machine Systems*, *44*(4), 429-440. http://dx.doi.org/ 10.1109/THMS.2014.2319822

Hauge, T.C., Katz, G.E., Davis, G.P., Huang, D-W., Reggia, J.A., & Gentili, R.J. (2017). High-level motor planning assessment during performance of complex action sequences. *In preparation.*

Heeringa, W. J. (2004). *Measuring dialect pronunciation differences using Levenshtein distance* (Doctoral dissertation, University Library Groningen [Host]).
Hinz, A. M., Kostov, A., Kneißl, F., Sürer, F., & Danek, A. (2009). A mathematical model and a computer tool for the Tower of Hanoi and Tower of London puzzles. *Information Sciences*, *179*(17), 2934-2947.
Hu, J. S., Lu, J., Tan, W. B., & Lomanto, D. (2015). Training improves laparoscopic tasks performance and decreases operator workload. *Surgical Endoscopy*. http://dx.doi.org/10.1007/s00464-015-4410-8
Hughes, C. (1996). Brief report: Planning problems in autism at the level of motor control. *Journal of autism and developmental disorders*, *26*(1), 99-107.
Jaquess, K. J., Gentili, R. J., Lo, L. C., Oh, H., Zhang, J., Rietschel, J. C., ... & Hatfield, B. D. (2017). Empirical evidence for the relationship between cognitive workload and attentional reserve. *International Journal of Psychophysiology*, *121*, 46-55.
Johnson-Frey, S., McCarty, M., & Keen, R. (2004). Reaching beyond spatial perception: Effects of intended future actions on visually guided prehension. *Visual cognition*, *11*(2-3), 371-399.
Kacer, J., Kutilek, P., Krivanek, V., Doskocil, R., Smrcka, P., & Krupka, Z. (2017, October). Measurement and Modelling of the Behavior of Military Pilots. In *International Conference on Modelling and Simulation for Autonomous Systems*(pp. 434-449). Springer, Cham.
Katz, G., Huang, D. W., Hauge, T., Gentili, R., & Reggia, J. (2017). A Novel Parsimonious Cause-Effect Reasoning Algorithm for Robot Imitation and Plan Recognition. *IEEE Transactions on Cognitive and Developmental Systems*.
Kawato, M. (1999). Internal models for motor control and trajectory planning. Current opinion in neurobiology, 9(6), 718-727.
Koenig, N., Takayama, L., & Matarić, M. (2010). Communication and knowledge sharing in human–robot interaction and learning from demonstration. *Neural Networks*, *23*(8), 1104-1112.
Kotovsky K, Hayes J.R., & Simon, H.A. Why are some problems hard? Evidence from tower of Hanoi. *Cognitive Psychology*, *17*, 248-294. https://doi.org/10.1016/0010-0285(85)90009-X.
Leusch, G., Ueffing, N., & Ney, H. (2003, September). A novel string-to-string distance measure with applications to machine translation evaluation. In *Proceedings of MT Summit IX* (pp. 240-247).
Levenshtein, V. I. (1966, February). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, No. 8, pp. 707-710).
Levine, S.J., & Williams, B.C. (2014). Concurrent Plan Recognition and Execution for Human-Robot Teams. *Proceedings of the Twenty-Fourth International Conference on Automated Planning and Scheduling (ICAPS)*, 490-498.
Magill, R., & Anderson, D. (2014). *Motor learning and control: Concepts and applications* (Tenth ed.). New York, NY: McGraw-Hill.
Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Keith Smith, J. E. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological review*, *95*(3), 340.
Meyer, M., van der Wel, R. P., & Hunnius, S. (2013). Higher-order action planning for individual and joint object manipulations. *Experimental brain research*, *225*(4), 579-588.

Miller, M. W., Rietschel, J. C., McDonald, C. G., & Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *International Journal of Psychophysiology*, *80*(1), 75-78. http://dx.doi.org/10.1016/j.ijpsycho.2011.02.003

Moran, D. W., & Schwartz, A. B. (1999). Motor cortical representation of speed and direction during reaching. *Journal of neurophysiology*, *82*(5), 2676-2692.

Naylor, J. C., & Briggs, G. E. (1963). Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of Experimental Psychology*, *65*(3), 217.

Nikolaidis, S., Hsu, D., & Srinivasa, S. (2017). Human-robot mutual adaptation in collaborative tasks: Models and experiments. *The International Journal of Robotics Research*, 0278364917690593.

Nikolaidis, S., Ramakrishnan, R., Gu, K., & Shah, J. (2015). Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*.

Noyes, J. M., & Garland, K. J. (2003). Solving the Tower of Hanoi: does mode of presentation matter?. *Computers in Human Behavior*, *19*(5), 579-592.

Ong, N. T., Lohse, K. R., & Hodges, N. J. (2015). Manipulating target size to influence perceptions of success when learning a dart-throwing skill. *Journal of Exercise, Movement, and Sport (SCAPPS refereed abstracts repository)*, *47*(1).

Ong, N. T., & Hodges, N. J. (2017). Balancing Our Perceptions of the Efficacy of Success-Based Feedback Manipulations on Motor Learning. *Journal of motor behavior*, 1-17.

Palmer, D. W. (1996). EXPLORING RECURSION WITH VARIATIONS ON THE TOWERS OF HANOI.

Paolo, A. M., Tröster, A. I., Axelrod, B. N., & Koller, W. C. (1995). Construct validity of the WCST in normal elderly and persons with Parkinson's disease. *Archives of Clinical Neuropsychology*, *10*(5), 463-473.

Paul, R. P., & Nof, S. Y. (1979). Work methods measurement—a comparison between robot and human task performance. *International Journal of Production Research*, *17*(3), 277-303.

Paxton, C., Jonathan, F., Hundt, A., Mutlu, B., & Hager, G. D. (2017). User Experience of the CoSTAR System for Instruction of Collaborative Robots. *arXiv preprint arXiv:1703.07890*.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: effects on visual search, discrimination, and decision making. *Journal of experimental psychology: Applied*, *9*(2), 119.

Rendell, M.A., Masters, R.S., Farrow, D., & Morris, T. (2011). An implicit basis for the retention benefits of random practice. *Journal of Motor Behavior. 43*(1), 1-13. http://dx.doi.org/10.1080/00222895.2010.530304

Rietschel, J. C., Miller, M. W., Gentili, R. J., Goodman, R. N., McDonald, C. G., & Hatfield, B. D. (2012). Cerebral-cortical networking and activation increase as a function of cognitive-motor task difficulty. *Biological Psychology*, *90*(2), 127-133. http://dx.doi.org/10.1016/j.biopsycho.2012.02.022

Rietschel, J. C., McDonald, C. G., Goodman, R. N., Miller, M. W., Jones-Lush, L. M., Wittenberg, G. F., & Hatfield, B. D. (2014). Psychophysiological support of increasing attentional reserve during the development of a motor skill. *Biological Psychology, 103,* 349-356. http://dx.doi.org/10.1016/j.biopsycho.2014.10.008

Rinehart, N. J., Bellgrove, M. A., Tonge, B. J., Brereton, A. V., Howells-Rankin, D., & Bradshaw, J. L. (2006). An examination of movement kinematics in young people with high-functioning autism and Asperger's disorder: further evidence for a motor planning deficit. *Journal of autism and developmental disorders*, *36*(6), 757-767.

Rosenbaum, D. A., Marchak, F., Barnes, H. J., Vaughan, J., Slotta, J. D., & Jorgensen, M. J. (1990). Constraints for action selection: Overhand versus underhand grips.

Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J., & van der Wel, R. (2012). Cognition, action, and object manipulation. *Psychological bulletin*, *138*(5), 924.

Ruiz-Rabelo, J. F., Navarro-Rodriguez, E., Di-Stasi, L. L., Diaz-Jimenez, N., Cabrera-Bermon, J., Diaz-Iglesias, C., . . . Briceño-Delgado, J. (2015). Validation of the NASA-TLX score in ongoing assessment of mental workload during a laparoscopic learning curve in bariatric surgery. *Obesity Surgery*, *25*, 2451-2456. http://dx.doi.org/10.1007/s11695-015-1922-1

Saemi, E., Porter, J. M., Ghotbi-Varzaneh, A., Zarghami, M., & Maleki, F. (2012). Knowledge of results after relatively good trials enhances self-efficacy and motor learning. *Psychology of Sport and Exercise*, *13*(4), 378-382.

Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1988). Procedural learning and neostyrial dysfunction in man. *Brain*, *111*(4), 941-960.

Salmerón, L., Gil, L., Bråten, I., & Strømsø, H. (2010). Comprehension effects of signalling relationships between documents in search engines. *Computers in Human Behavior*, *26*(3), 419-426.

Salter, T., Dautenhahn, K., & te Boekhorst, R. (2006). Learning about natural human–robot interaction styles. *Robotics and Autonomous Systems*, *54*(2), 127-134.

Sanford, J., Moreland, J., Swanson, L. R., Stratford, P. W., & Gowland, C. (1993). Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. *Physical therapy*, *73*(7), 447-454.

Schiff, R., & Vakil, E. (2015). Age differences in cognitive skill learning, retention and transfer: The case of the Tower of Hanoi Puzzle. *Learning and Individual Differences, 39*, 164-171.

Scott, S. H., Gribble, P. L., Graham, K. M., & Cabel, D. W. (2001). Dissociation between hand motion and population vectors from neural activity in motor cortex. *Nature*, *413*(6852), 161-165.

Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, *5*(7), 532-546.

Sergio, L. E., & Kalaska, J. F. (2003). Systematic changes in motor cortex cell activity with arm posture during directional isometric force generation. Journal of Neurophysiology, 89(1), 212-228. Chicago.

Shah, J., & Breazeal, C. (2010). An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. *Human factors*, *52*(2), 234-245.

Shah, J., Wiken, J., Williams, B., & Breazeal, C. (2011, March). Improved human-robot team performance using chaski, a human-inspired plan execution system. In *Proceedings of the 6th international conference on Human-robot interaction* (pp. 29-36). ACM.

Shadmehr, R., & Wise, S. P. (2005). *The computational neurobiology of reaching and pointing: A foundation for motor learning*. Cambridge, MA: MIT Press.

Scholtz, J. (2003). Theory and evaluation of human robot interactions. *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*, 2003, pp. 10 doi: 10.1109/HICSS.2003.1174284.

Scholtz, J., Young, J., Drury, J. L., & Yanco, H. A. (2004, April). Evaluation of human-robot interaction awareness in search and rescue. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on* (Vol. 3, pp. 2327-2332). IEEE.

Shuggi, I.M., Oh, H., Shewokis, P.A., & Gentili, R.J. (2017a). Mental workload and motor performance dynamics during practice of reaching movements under various levels of task difficulty. *Neuroscience*, *360,* 166-179. https://doi.org/10.1016/j.neuroscience.2017.07.048.

Shuggi, I. M., Shewokis, P. A., Herrmann, J. W., & Gentili, R. J. (2017b). Changes in motor performance and mental workload during practice of reaching movements: a team dynamics perspective. *Experimental brain research*, *236*(2), 433-451.

Shute, G. E., & Huertas, V. (1990). Developmental variability in frontal lobe function. *Developmental Neuropsychology*, *6*, 1-11.

Sober, S. J., & Sabes, P. N. (2003). Multisensory integration during motor planning. *Journal of Neuroscience*, *23*(18), 6982-6992.

Sober, S. J., & Sabes, P. N. (2005). Flexible strategies for sensory integration during motor planning. *Nature neuroscience*, *8*(4), 490-497.

Soukoreff, R. W., & MacKenzie, I. S. (2001, March). Measuring errors in text entry tasks: an application of the Levenshtein string distance statistic. In *CHI'01 extended abstracts on Human factors in computing systems* (pp. 319-320). ACM.

Steenbergen, B., & Gordon, A. M. (2006). Activity limitation in hemiplegic cerebral palsy: evidence for disorders in motor planning. *Developmental medicine and child neurology*, *48*(9), 780-783.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006, March). Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 33-40). ACM.

Ste-Marie, D. M., Vertes, K. A., Law, B., & Rymal, A. M. (2013). Learner-controlled self-observation is advantageous for motor skill acquisition. *Frontiers in Psychology*, *3*(556) http://dx.doi.org/10.3389/fpsyg.2012.00556.

Stevens, D., Anderson, D. I., O'Dwyer, N. J., & Williams, A. M. (2012). Does self-efficacy mediate transfer effects in the learning of easy and difficult motor skills? *Consciousness and Cognition*, 21(3), 1122-1128. http://dx.doi.org/10.1016/j.concog.2012.03.014

Taig, E., Küper, M., Theysohn, N., Timmann, D., & Donchin, O. (2012). Deficient use of visual information in estimating hand position in cerebellar patients. *Journal of Neuroscience*, *32*(46), 16274-16284.

Tenorth, M., Ziegltrum, J., & Beetz, M. (2013, November). Automated alignment of specifications of everyday manipulation tasks. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on* (pp. 5923-5928). IEEE.

Thibaut, J.-P., & Toussaint, L. (2010). Developing motor planning overages. *Journal of Experimental Child Psychology, 105*, 116–129. doi: 10.1016/j.jecp.2009.10.003

Trempe, M., Sabourin, M., & Proteau, L. (2012). Success modulates consolidation of a visuomotor adaptation task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 52.

Uno, Y., Kawato, M., & Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement. *Biological cybernetics*, *61*(2), 89-101.

Vakil, E., & Heled, E. (2016). The effect of constant versus varied training on transfer in a cognitive skill learning task: The case of the Tower of Hanoi Puzzle. *Learning and Individual Differences*, *47*, 207-214.

Vink, P., Peeters, M., Gründemann, R. W. M., Smulders, P. G. W., Kompier, M. A. J., & Dul, J. (1995). A participatory ergonomics approach to reduce mental and physical workload. *International Journal of Industrial Ergonomics*, *15*(5), 389-396.

Wagner, R. A., & Fischer, M. J. (1974). The string-to-string correction problem. *Journal of the ACM (JACM)*, *21*(1), 168-173.

Wang, N., Pynadath, D., & Hill, S. (2015). Building Trust in a Human Robot Team with Automatically Generated Explanations. I. *ITSEC Paper*, (15315), 1-12.

Welsh, M. C. (1991). Rule-guided behavior and self-monitoring on the Tower of Hanoi disk-transfer task. *Cognitive Development*, *6*(1), 59-76.

Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain and cognition*, *41*(2), 231-242.

Welsh, M. C., & Huizinga, M. (2005). Tower of Hanoi disk-transfer task: Influences of strategy knowledge and learning on performance. *Learning and Individual Differences*, *15*(4), 283-298.

Wickens, C., Kramer, A., Vanasse, L., Donchin, E. (1983). Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. *Science*, *221*, 1080–1082.

Wilmut, K., Byrne, M., & Barnett, A. L. (2013). To throw or to place: does onward intention affect how a child reaches for an object?. *Experimental brain research*, *226*(3), 421-429.

Wilson, G.F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *International Journal of Aviation Psychology*, *12*(1), 3–18. http://dx.doi.org/10.1207/S15327108IJAP1201_2

Wilson, G.F., & Russell, C.A. (2003). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors*, *45*(4), 635-643.

Wunsch, K., Pfister, R., Henning, A., Aschersleben, G., & Weigelt, M. (2016). No interrelation of motor planning and executive functions across young ages. *Frontiers in psychology*, *7*, 1031.

Young, S. J., Pratt, J., & Chau, T. (2009). Target-directed movements at a comfortable pace: Movement duration and Fitts's law. *Journal of Motor Behavior*, *41*(4), 339-349. http://dx.doi.org/10.3200/JMBR.41.4.339-346

Yujian, L., & Bo, L. (2007). A normalized Levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence*, *29*(6), 1091-1095.

Zhang, Y., Narayanan, V., Chakraborti, T., & Kambhampati, S. (2015). A human factors analysis of proactive support in human-robot teaming. *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on. IEEE, 2015*. https://doi.org/ 10.1109/IROS.2015.7353878.