THE UNIVERSITY *of* EDINBURGH

Advancements in Sensory-Motor Perception and Biologically-Inspired Hierarchical Learning for Embodied Intelligence

Eleftherios Triantafyllidis

Doctor of Philosophy

Institute of Perception, Action and Behaviour
School of Informatics
University of Edinburgh

2024

Abstract

From a biological perspective, humans possess incredible sensory, dexterous and cognitive abilities. By virtue of these abilities, humans and more broadly their sensory systems are able to adapt to environmental demands seamlessly. Switching from a biological perspective to robotic systems as embodied intelligence, achieving such adaptation is currently far from trivial. Understanding the biological mechanisms that govern and render humans proficient in interacting with their surroundings, could ultimately illuminate new pathways to replicate such human-like cognition and dexterity in machines. Inspired by the aforementioned narrative, this thesis delves into and addresses four main research areas.

The first contribution of this thesis provides novel insights into the intricate interplay of multimodal interfaces, their impact on the human sensory-motor system and their correlation to the generation of meaningful motor actions. Different sensory modalities are examined, entailing a full factorial comparison of auditory, visual and somatosensory states and their influence on motor performance. Through a series of varying complexity motor tasks with human subjects, a correlation is established between sensory states and their influence on motor actions. Results provide novel evidence of which sensory combinations contribute to enhanced task performance and how these can be harnessed.

The second contribution of this work is the derivation of a novel metric capable of quantifying motor actions stemming from the intricate human sensory-motor system. Measuring human motor performance is a complex phenomenon and the absence of a standardised metric renders inter-study comparability challenging. To this end, four motor tasks, increasing in spatial complexity, were devised to establish a correlation of which spatial variables influence motor performance. Results revealed which spatial variables had the most notable effect, highlighting that existing metrics are inadequate for modelling higher dimensions. To account for this, a novel metric is derived, capable of modelling human motor performance in full 3D space, underlining its value for quantifying commonly seen motor movements and enhancing inter-study comparability.

The third and penultimate contribution builds on the foundation laid by the preceding segments, striving to align human capabilities closer to embodied intelligence. To realise this aim, inspired by a biological standpoint, the RObotic MAnipulation Network (ROMAN) is introduced. ROMAN is a novel Hybrid Hierarchical Learning (HHL) architecture designed to address the challenges of notably complex long-horizon sequential tasks.

ROMAN utilises the exploratory nature of Reinforcement Learning (RL) while simultaneously exploiting the higher-level skills of humans in the form of imitation. Consisting of a plethora of specialising skills, ROMAN's hierarchical architecture demonstrates versatility in intricate, long-horizon sequential tasks; while exhibiting robustness against various levels of sensory uncertainties. By virtue of the HHL employed, ROMAN also exhibits adaptability beyond demonstrated behaviour; featuring failure recovery capabilities and adaptation in avoiding local minima. These results underline the significance of ROMAN for autonomous manipulation tasks necessitating intelligent and adaptive behaviour.

The fourth and concluding contribution of this thesis investigates the potential of language-guided exploration in augmenting embodied intelligence. In pursuit of this goal, the Intrinsically Guided Exploration from Large Language Models (IGE-LLMs) framework is presented, capable of complementing the existing bio-inspired hierarchy of ROMAN. By harnessing LLMs as an assistive intrinsic reward source alongside the conventional RL paradigm, IGE-LLMs enhances the exploratory process to address intricate settings challenged by sparse rewards and long-horizons. Validated on environments challenged by exploration and long-horizons, IGE-LLMs exhibits notably higher performance over existing methods and is capable of complementing the shortcomings of using LLMs in isolation. Moreover, the modularity and robustness of IGE-LLMs is underscored, due to its ability to complement existing intrinsic reward methods and its insensitivity to most intrinsic scaling parameters. Finally, the framework's resilience is highlighted over existing methods when faced with increased uncertainties and horizons. Capable of fostering exploration and the ability to automate the orchestration of ROMAN's intricate macro-actions, IGE-LLMs value as a language-guided framework is underlined.

This thesis provides novel findings on harnessing human sensory-motor abilities for generating meaningful motor actions, which can be adequately measured and quantified. Ultimately, these represent the inspiration for shaping the development of a novel bio-inspired learning method to align human capabilities closer to embodied intelligence that can further be complemented and automated by eliciting language-guided exploration; tailored to address notably intricate, long-horizon tasks with sparse rewards. Nevertheless, to further narrow the gap between humans and machines, a deeper understanding of designing artificial intelligence inspired by biological insights is necessitated.

Lay Summary

Humans have a remarkable ability to interact with their surroundings. With years of evolution, the intricate sensory-motor system has been honed to unmatched degrees, allowing humans to meet environmental demands with their surroundings with astonishing efficiency and accuracy. Now switching from a biological perspective to robots as embodied intelligence, meeting environmental demands with such efficiency is currently far from trivial. This research aims to uncover the mysteries behind the capabilities exhibited by humans to align these closer to embodied intelligence with the goal of enhancing their overall skills. To achieve the aforementioned, this thesis delves into four in total areas of research.

The first part of this research explored how humans interact with their surrounding environment and in particular how they utilise the plethora of different senses they possess. With this aim in view, this thesis delves into how different sensory modalities affect human motor performance in daily manipulation tasks. In particular, emphasis is given to how these sensory cues enable the human sensory-motor system to generate meaningful motor actions. This, will in turn shed light on the underlying biological mechanisms that render humans so proficient when meeting intricate environmental demands. Consequently, to achieve this, the first part of this research looks at how combining different senses, such as hearing, sight, and touch, affects motor performance across a variety of manipulation tasks in a 3D environment. By recruiting different human participants, an evaluation of these sensory cues and their combinations is performed. To this end, novel findings are presented and a correlation is established as to which sensory cues in a multimodal interface, correspond to improved task performance across different measurements both subjective and objective. These findings suggest the value of using multimodal interfaces in settings where human operation or intervention is of necessity, with the knowledge of prioritising certain combinations of sensory states, to harness and maximise sensory-motor performance.

Having a better understanding of how the sensory system in humans contributes to generating meaningful motor actions, it is now imperative to be able to measure and quantify such actions. A notable challenge amongst current work is the lack of a universal and standardised metric to evaluate human motor performance and most notably 3D movements and actions. This, in turn, renders comparisons between different studies notably difficult. The second part of this research project delves into existing human motor per-

formance metrics. After reviewing the related work, it was identified that current metrics are mostly limited to lower-dimensional movements. With the aim of deriving a higher-dimensional metric utilising well-established formulations as a basis, a user study was conducted whereby participants engaged in four increasingly more complex motor tasks to understand which factors influenced 3D human-motor movement the most. Detailed statistical analysis indicated which spatial variables played pivotal roles in performance. It also became evident that many existing metrics are unable to fully capture the nuances of complex human motor movements in expansive 3D spaces. Addressing this, the second part of this thesis introduced a new comprehensive metric for evaluating human motor movements in 3D space entailing high degrees of freedom. This novel formulation holds value for quantifying upper-body motor movements and interactions in virtual reality and robotic-controlled tasks. This, in turn, facilitates more consistent comparisons between different studies in the broader realm of human ergonomics research with generalising capabilities.

From the preceding research, a better understanding of the human sensory-motor system was revealed such that human-like capabilities can ultimately be aligned closer to embodied intelligence. Inspired by a biological standpoint, the penultimate aim and third part of this thesis delved into the workings of Artificial Intelligence (AI), introducing the RObotic MAnipulation Network (ROMAN). ROMAN is a novel, hybrid hierarchical learning architecture tailored for embodied intelligence faced with intricate challenges in long-duration robotic manipulation tasks entailing a plethora of sequences. ROMAN combines the adventurous learning and exploratory nature of reinforcement learning, fusing these with human-like capabilities in the form of imitation skills. At its core, ROMAN is built of a plethora of expert components, all overseen and supervised by a central gating network known as the manipulation network. The hierarchical architecture, hybrid learning and high-level task decomposition are key features of ROMAN. By virtue of these features, much like observed in biology, ROMAN is tailored for intricate tasks, whereby even under sensory uncertainty and deviation from sensory readings, robust performance can be maintained even amongst the longest in duration tasks. Most notably, ROMAN is capable of adapting and going beyond imitated human behaviour, due to its ability to recover outside of demonstrated cases and dynamically adapt to environmental demands. These attributes solidify ROMAN as a pivotal system for complex robotic manipulation tasks requiring intelligent and adaptive behaviour.

The final part of this thesis aims to further solidify AI and to further narrow the gap between biology and machines. With this objective in mind, this research delves into the workings of the power of language. Inspired again from a biological standpoint, an emphasis on language as a tool is given. This is much like how the human cognitive system is enriched surrounding the think-aloud thought process, to elicit meaningful reasoning for complex tasks. To this end, the final research of this work delves into the usage of Large Language Models (LLMs), deriving a novel framework – the Intrinsically Guided Exploration from Large Language Models (IGE-LLMs). By leveraging the context and common-sense aware reasoning of LLMs at its core, IGE-LLMs is tailored for notably complex, of long-duration tasks, whereby environmental feedback in the form of rewards may not be readily apparent. Results show that IGE-LLMs outperform related learning approaches that utilise LLMs, overcoming their inherent limitations when used as main policy drivers. Moreover, further evaluation reveals that IGE-LLMs exhibit stable and robust performance even under increasing complexities. Additionally, IGE-LLMs modularity is underscored due to the framework's ability to be combined and complement existing learning paradigms and extend ROMAN's architecture. Overall, these results underline the value of IGE-LLMs as a promising basis for guiding intelligent agents to excel in complex tasks, by providing valuable language-guided exploratory potential.

In summary, this thesis provides novel insights into enhancing, harnessing and quantifying the human sensory-motor system, to elicit the underlying biological mechanisms that govern this highly advanced system. These insights are then utilised as a blueprint to align human-like capabilities closer to embodied intelligence and most notably to robots, to address notable challenges in manipulation and environmental interactions. Ultimately, this led to the derivation of two novel frameworks, ROMAN and IGE-LLMs, aiming to align embodied intelligence closer to the capabilities of humans and more broadly biology. Nevertheless, to align machines closer to human-like functioning, additional research is necessitated to address the profound need for further understanding of how to craft embodied intelligence based on biological principles.

Acknowledgements

Foremost, I would like to thank my primary advisor Dr. Zhibin (Alex) Li for his overall support during my studies both in regards to research and professionalism; introducing me to the exciting world of robotics and sharing his experience and knowledge throughout. His trust in my skills, allowed me to pursue my passion for research and thanks to his unwavering support, guidance and insightful feedback, I found both enjoyment and excitement in my studies and academic research pursuits.

I also thank my second advisor Prof. Taku Komura for his invaluable support and guidance during my academic research and career development. I would also like to thank Prof. Robert Fisher for his annual feedback regarding the research project.

I would also like to thank all of my colleagues, some of whom I call friends at the University, the Institute of Perception, Action and Behaviour and the Edinburgh Centre for Robotics. Their constant support, guidance and feedback throughout these years have undoubtedly made the pursuit of this doctorate a pleasant and exciting experience. In particular, I would like to thank Iordanis, Chris, Wenbin, Fernando, Zhaocheng, Wouter, Quentin, Wanming, Ruoshi, Kai, and Chanyu from my group at the Advanced Intelligent Robotics (AIR) Lab as well as Filippos and Giorgos from my cohort for their help and input.

I would moreover like to thank both the internal examiner Dr. Xiaoxuan Chris Lu and external examiner Prof. Sylvain Calinon for their invaluable and constructive feedback as well as rigorous review regarding the doctoral thesis.

This thesis would not have been possible without the overall support of my family and allowing me to pursue my dreams and ambitions. Their love and support have made this very thesis possible. I will always be thankful for their support and their trust in me and their efforts throughout these years have made me the person I am today.

I would especially like to thank my partner Valentina for her incredible support during my doctoral studies over all these last years without whom I would not have the strength and inspiration needed for this journey. I am truly thankful to have her in my life.

I would also like to thank all my friends all these years who stood by me and continuously provided their support and friendship. I am proud to have such people in my life.

Lastly, a special thanks to Jens Angerer for his trust in my skills during my professional development and for introducing me to the exciting world of mixed reality.

Funding Support

This research was supported by the Engineering and Physical Sciences Research Council (EPSRC) as part of the Centre for Doctoral Training in Robotics and Autonomous Systems at Heriot-Watt University and The University of Edinburgh, *(Grant Number: EP/L016834/1)*, the EPSRC Future AI and Robotics for Space *(Grant Reference and Number: 10.13039/501100000266 and EP/R026092/1)*, the FairSpace project *(Grant Number: EP/R026092/1)* as well as the Horizon 2020 project Harmony *(Grant Number: 101017008)*.

Ethics Approval

All work herein presented concerned with the recruitment of human participants, Chapter 3 and Chapter 4, were approved by the ethics committee from the School of Informatics at the University of Edinburgh. Ethics approval reference code(-s): 2019/49550 and 2019/22258.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Eleftherios Triantafyllidis*)

Για την οικογένεια μου, τη σύντροφο μου, τους φίλους μου,
και όλα τα άτομα που με στήριξαν στη ζωή μου.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AI Artificial Intelligence. 1, 5, 6, 8, 9, 16, 18, 46, 52, 132, 171, 193, 194, 198, 201, 202

AMOLED Active-Matrix Organic Light-Emitting Diode. 66

ANOVA ANalysis Of VAriance. 73

API Application Programming Interface. 178

ART Aligned Rank Transform. 73, 74, 76, 77, 104, 109, 110, 112

BC Behavioural Cloning. 48, 49, 51, 55, 129, 135–140, 143, 160–163, 165, 181, 205, 209, 211, 218, 219, 223

CI Confidence Intervals. 41, 42, 44, 73, 104

CLIP Contrastive Language-Image Pre-Training. 202

CoT Chain of Thought Reasoning. 186, 188–190, 202

CPU Central Processing Unit. 204

DC Direct Current. 67

DoF Degree of Freedom. 4, 7, 11, 29, 30, 37, 38, 58, 74, 86, 88–90, 93, 96, 97, 99, 100, 111, 113, 121, 126, 194, 196, 199

DZ Dead Zone. 149

FoV Field of View. 22, 23, 63, 66, 94

FPV First Person View. 23

GAIL Generative Adversarial Imitation Learning. 49, 55, 129, 135–140, 143, 160, 161, 163, 165, 181, 205, 209, 211, 218, 219, 222–225

GAN Generative Adversarial Network. 49

GGC Greenhouse–Geisser Correction. 73, 74, 76, 80, 104

GPU Graphics Processing Unit. 63

HCI Human-Computer Interaction. 14, 29, 30, 92, 121

HHL Hybrid-Hierarchical Learning. 8, 11, 129–131, 133, 139, 145, 148, 149, 151, 160–162, 164, 166, 169, 170, 179, 195–198, 200, 209, 218, 219

HIL Hierarchical Imitation Learning. 51, 160, 161

HL Hierarchical Learning. 50

HRI Human-Robot Interaction. 121

HRL Hierarchical Reinforcement Learning. 50, 51, 160, 161

ICM Intrinsic Curiosity Module. 54, 177, 185, 191

ID Index of Difficulty. 30

IGE-LLMs Intrinsically Guided Exploration from Large Language Models. 9, 10, 12, 15, 16, 172, 173, 175, 176, 178–183, 185–187, 189–192, 194, 195, 197, 198, 201, 202, 226

IK Inverse Kinematic. 134, 204

IL Imitation Learning. 47–51, 55, 135, 138, 160, 161, 169, 218, 219

IMU Inertial Measurement Unit. 68

IPD Inter-Pupillary Distance. 72, 103

IPS In-Plane Switching. 66

IRL Inverse Reinforcement Learning. 49

LfD Learning from Demonstration. 47, 48

LLM Large Language Model. 5, 9, 10, 12, 14–16, 18, 52, 55, 56, 171–173, 175, 176, 178–182, 184–190, 192, 195, 197, 198, 201, 202, 226, 227

LMHC Leap Motion Hand Controller. 63, 66, 68, 94, 96

MDP Markov Decision Process. 176

ML Machine Learning. 46

MLP Multi-Layer Perceptron. 141, 222

MN Manipulation Network. 8, 11, 129, 130, 133, 134, 139–143, 145, 148–154, 157–159, 163–170, 180, 181, 183, 184, 190–192, 197, 198, 201, 204–217, 220–225

MoE Mixture of Experts. 8, 50, 51, 130

MR Mixed Reality. 3, 11, 18, 28–31, 35, 38, 44, 57–59, 74, 88, 90, 91, 93, 100, 109, 121, 143, 170, 193, 196, 200, 201

MSE Mean Squared Error. 135, 139

MT Movement Time. 30

NASA-TLX National Aeronautics and Space Administration - Task Load Index. 71, 72, 74–77, 82, 83

NN Neural Network. 49, 50, 130, 133–136, 138–143, 145, 148–152, 154–159, 166, 170, 177, 181, 182, 184, 201, 203–212, 214, 216, 217, 221–225

OI Object of Interest. 142, 160, 185

PCA Principal Component Analysis. 167

PD Proportional Derivative. 64, 96

PPO Proximal Policy Optimization. 47, 135, 138–140, 143, 160, 161, 177, 179, 181, 211, 218, 219, 223, 225

RGB Red Green and Blue. 140, 158

RL Reinforcement Learning. 8–10, 12, 15, 18, 46–53, 56, 129, 132, 134–139, 149, 160–163, 165, 166, 169, 172–176, 178, 179, 181, 182, 185, 187, 189–191, 197, 198, 204, 205, 209, 218, 219

RM-ANOVA Repeated Measures - ANalysis Of VAriance. 73, 74, 76, 77, 80, 104, 106–110, 112, 114, 115

RND Random Network Distillation. 54, 178, 185, 191

ROMAN RObotic MAnipulation Network. 8–12, 15, 129–135, 137–143, 145–174, 176, 177, 179–184, 186–192, 194, 196–198, 200, 201, 203–227

ROS Robot Operating System. 96, 134, 204

RPM Revolutions Per Minute. 67

SAC Soft Actor-Critic. 47, 179, 180

SDK Software Development Kit. 63, 94

SHAP Southampton Hand Assessment Procedure. 43, 68–70, 200

SUS System Usability Scale. 71, 72, 75–77

t-SNE t-Distributed Stochastic Neighbor Embedding. 153, 160, 165, 167, 168, 201

URDF Unified Robot Description Format. 96, 97, 134, 204

VE Virtual Environment. 18, 24, 30–32, 38, 44, 57, 88, 91, 92, 94, 96, 99, 100, 120, 121

VR Virtual Reality. 18, 25, 37, 41, 43, 96, 103, 118, 122, 127, 170

VRHMD Virtual Reality Head Mounted Display. 22, 25, 35, 61–64, 66, 72–74, 76, 77, 79–84, 86, 87, 94, 96, 103, 195, 199

List of Symbols

MT	movement time, as a line equation based on a, b and ID
a	intercept constant (determined via regression)
b	slope constant (determined via regression)
ID	index of difficulty, as a logarithmic ratio (log_2) of A and B
A	target distance
W	target width
F	probe or index finger size
θ	inclinational or polar angle
ϕ	directional or azimuth angle
α	rotational distance to target
ω	rotational tolerance
r^2, R^2	coefficient of determination
$\vec{b_i}$	parent joint
$\vec{b_{i-1}}$	child joint
$u(t)$	angular velocity control signal via a PD controller
$e(t)$	positional error value
$q_h(t)$	human joint position
$q_r(t)$	robot joint position
$\dot{e(t)}$	velocity error value
$\dot{q}_{ref}$	reference joint velocity
$\dot{q}_r(t)$	robot joint velocity
K_P	proportional gain

K_D	derivative gain
K_E	kinetic energy
m	mass
v	velocity
P	virtual relative penetration value
v_r	robot segment, utilised by P
v_o	object segment, utilised by P
s_o	centre to surface distance of an object, utilised by P
V	vibration intensity
L, W, H	length, width and height
ϵ	estimate of sphericity, statistics
p	p-value for significance as a probability, statistics
F	F-test, statistics
η_p^2	measure of effect size, statistics
χ^2	chi-squared test, statistics
τ	torque
θ_d	desired angle
θ_c	current angle
$\dot{\theta}$	velocity of virtual hand
π	policy
s_t	state vector at time step t
a_t	action space at time step t
τ_A	agent trajectories
τ_E	human expert trajectories
∇	gradient
G	generator
D	discriminator / dictionary (context-dependent)
θ	weight or policy network parameters (context-dependent)
π_θ	policy parameterised by weights θ
k	step

$L(s, a, \theta_k, \theta)$	loss function for PPO
r^i	intrinsic reward
r^e	extrinsic reward
r, r^c	total reward, in this case, denoted as the sum of r^i and r^e
λ^i	scaling parameter for intrinsic reward r^i
λ^e	scaling parameter for extrinsic reward r^e
x_g	gripper action
w	weights
σ	normalised exponential function
z	input vector
e	standard exponential function
K	number of classes in the multi-class classifier
$\sigma(\pm)$	standard deviation
S	state space
A	action space
$P(s'\|s, a)$	transition function, probability of transitioning to next state
$R(s, a, s')$	extrinsic reward function
γ	discount factor
$\pi(a\|s)$	policy defining probability of taking action a in state s
r_t^i	intrinsic reward at time-step t
$N(s_t)$	count of the encountered current state
α	scaling factor (ICM)
$\widehat{\phi}$	predicted representation of the next state given the current state and action
ϕ	actual representation of the next state
r^i	intrinsic reward
$f : S \times A \to [0, 1]$	function evaluating potential future rewards
$D : S \times A \to [0, 1]$	dictionary representing the evaluation of potential future rewards
r^c	total combined reward
λ^i	intrinsic scaling factor
w^i	linearly decaying weight for the intrinsic reward

List of Publications

First Authored Publications

The research presented in this thesis led to the following first authored publications:

1. **E. Triantafyllidis**, F. Christianos and Z. Li. (2023). "Intrinsic Language-Guided Exploration for Complex Long-Horizon Robotic Manipulation Tasks". In: *International Conference on Robotics and Automation (ICRA 2024)*.

2. **E. Triantafyllidis**, F. Acero, Z. Liu and Z. Li. (2023). "Hybrid hierarchical learning for solving complex sequential tasks using the robotic manipulation network ROMAN". In: *Nature Machine Intelligence*, Volume: 5, Issue: 9, Pages: 991–1005, DOI: 10.1038/s42256-023-00709-2.

3. **E. Triantafyllidis**, C. Mcgreavy, J. Gu and Z. Li. (2020). "Study of Multimodal Interfaces and the Improvements on Teleoperation". In: *IEEE Access*, vol. 8, pp. 78213-78227, DOI: 10.1109/ACCESS.2020.2990080.

4. **E. Triantafyllidis**, C. Yang, C. McGreavy, W. Hu and Z. Li. (2020). "Robot intelligence for real-world applications". In: *AI for Emerging Verticals: Human-robot computing, sensing and networking, IET Computing and Networks*, Book Chapter, Chapter 4.0.

5. **E. Triantafyllidis**, W. Hu, C. McGreavy and Z. Li. (2021). "Metrics for 3D Object Pointing and Manipulation in Virtual Reality: The Introduction and Validation of a Novel Approach in Measuring Human Performance". In: *IEEE Robotics & Automation Magazine*, DOI: 10.1109/MRA.2021.3090070.

6. **E. Triantafyllidis** and Z. Li. (2021). "The Challenges in Modeling Human Performance in 3D Space with Fitts' Law". In: *CHI Conference on Human Factors in*

Computing Systems Extended Abstracts (CHI '21 Extended Abstracts). Association for Computing Machinery, May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA. DOI: 10.1145/3411763.3443442.

7. **E. Triantafyllidis** and Z. Li. (2021). "Considerations and Challenges of Measuring Operator Performance in Telepresence and Teleoperation Entailing Mixed Reality Technologies". In: *CHI Conference on Human Factors in Computing Systems Workshop CHI '21 (Evaluating User Experiences in Mixed Reality)*. Association for Computing Machinery, May 7, 2021, Yokohama, Japan. ACM, New York, NY, USA.

Publications as Co-Author

Parts of this research thesis contributed to the following co-authored publications:

1. W. Yu, C. Yang, C. McGreavy, **E. Triantafyllidis**, G. Bellegarda, M. Shafiee, A. J. Ijspeert and Z. Li. (2023). "Identifying Important Sensory Feedback for Learning Locomotion Skills". In: *Nature Machine Intelligence*, Volume: 5, Issue: 8, Pages: 919–932, DOI: 10.1038/s42256-023-00701-w.

2. W. Hu, F. Acero, **E. Triantafyllidis**, Z. Liu and Zhibin Li. (2023). "Modular Neural Network Policies for Learning In-flight Object Catching with a Robot Hand-Arm System". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Chapter 1

Introduction

At the intersection of biology and Artificial Intelligence (AI) lies a compelling aspiration: the quest for biologically-inspired embodied intelligence. From a biological standpoint, humans are able to navigate in intricate environments, meet complex environmental demands and generally perform motor tasks entailing a plethora of sequences that span over long horizons seamlessly. The exceptional ability of humans to achieve such levels of physical interactions with their surroundings is not just a product of evolution but a testament to the profound intricacies of human perception, cognition, and motor control.

Switching from an AI perspective as robots with embodied intelligence, achieving such interactions is currently far from trivial and stands in stark contrast to the capabilities exhibited by humans. Current embodied intelligence struggles in intricate settings especially when these entail a plethora of different sequences that span over a long horizon whereby immediate feedback from the environment is not readily apparent. Naturally, this raises an interesting question: Can machines be designed and taught in such a manner that one day these systems can achieve a similar level of embodied intelligence rooted in biological inspiration?

The duality of the exploration necessitated to answer this question becomes increasingly apparent. To journey towards this vision, it is imperative to first understand the biological mechanisms that govern the abilities exhibited by humans that render them proficient in interacting with their surroundings, all with the hopes of replicating such behaviour in machines. The first milestone towards this endeavour is to understand the very fabric

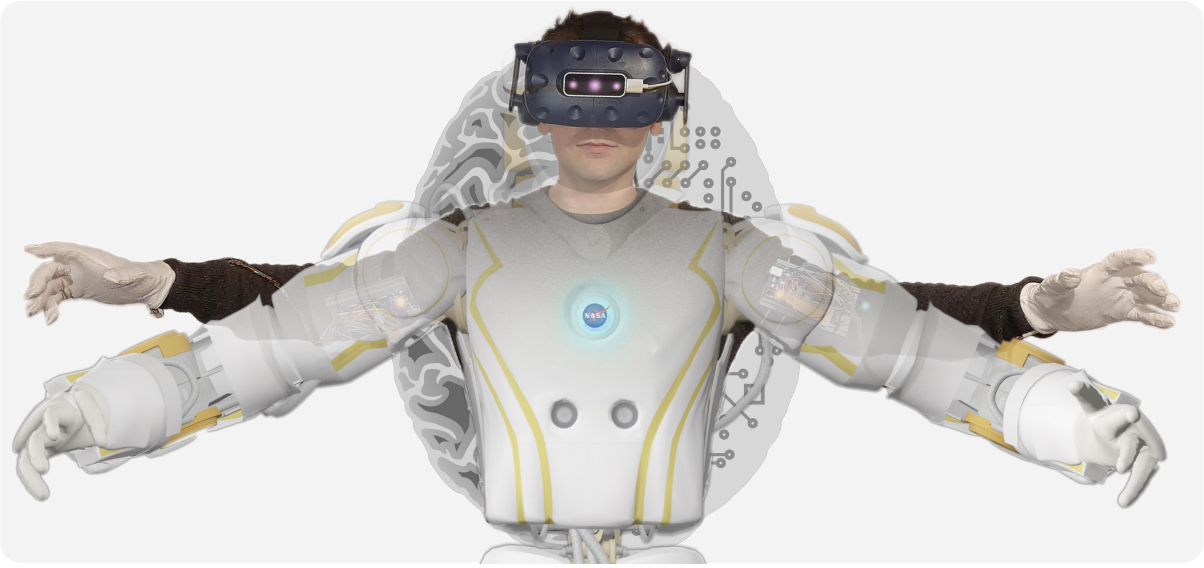

Figure 1.1: **Artistic illustration of the future synergy that could exist between humans and machines to create biologically-inspired embodied intelligence.**

of the human sensory-motor perceptive system. The central question guiding this system revolves around how the human brain integrates various sensory inputs, coordinates the body's motor responses and processes complex information. In particular, what innate factors render the human sensory system superior and how do diverse sensory inputs translate into purposeful motor actions that allow humans to tackle intricate environmental challenges with seemingly minimal effort? Given this knowledge, how can meaningful embodied intelligence methods be derived to solve intricate manipulation settings in robotics?

This thesis embarks on this journey. The first half of this thesis delves into understanding the human sensory-motor system and how it can be elucidated, harnessed and ultimately measured. Thereafter, given this understanding as a blueprint, the remainder and larger part of this research will aim to advance biologically inspired embodied intelligence in robotics to address intricate settings.

1.1 Problem Formulation and Research Questions

To effectively align embodied intelligence more closely with biology and the notable capabilities inherent in humans, **four** central questions and challenges must be addressed. Figure 1.1 visually depicts how the fusion between biology and machines may look like.

Firstly, an important question surrounding the human sensory-motor system needs to be answered. *What are the underlying factors that render the human sensory system so effective, and how does the human brain, in synergy with the broader sensory system, integrate a plethora of sensory inputs to initiate meaningful motor actions?* To answer these questions, it is initially important to understand how the human sensory-motor system works and how it can be harnessed. Humans utilise a wide range of stimuli to generate meaningful motor actions to meet everyday environmental demands (Heilig, 1992). It is widely considered that motor actions are inherently multimodal, necessitating the simultaneous presence of two or more modalities as supported by Billard and Kragic (2019). Neuroimaging studies have further solidified these claims, showing that visual and somatosensory stimulation overlap in the same region of the brain, the occipital lobe according to A Ghazanfar and E Schroeder (2006); Amedi et al. (2001); James et al. (2002). It would thus come natural to understand how different types of sensory stimulation affect the sensory system and more broadly the sensory-motor performance in humans. To this end and as part of the first investigation of this thesis, an emphasis is given on multimodal interfaces entailing different types of sensory stimulation and how these affect motor performance. The research outcome of this investigation would be two-fold. First, unravelling the underlying sensory cues that affect motor performance is expected to be invaluable for Mixed Reality (MR) and robotic teleoperation-based scenarios, whereby human operators are an inseparable part of the decision-making process (Sigrist et al., 2013; Triantafyllidis et al., 2020; Lipton et al., 2018). Secondly, with the hopes of ultimately illuminating pathways to replicate human-like cognition and dexterity to embodied intelligence, it is crucial to understand how the human sensory system can be harnessed and its correlation to motor proficiency (Triantafyllidis et al., 2023b).

Secondly, while unravelling the complexities of the human sensory-motor system is the first milestone, a subsequent challenge of quantifying this system arises. *How can such phenomena be quantified, translating organic experiences into measurable metrics?* Imagining that the human sensory-motor system can be harnessed and the underlying correlation between sensory cues and motor-action performance understood, it is now crucial to measure this behaviour. In the broader context of evaluating emerging methods and more specifically those aiming to understand human behaviour and more notably the sensory-motor system, different evaluation metrics are necessitated (Gonzalez-Franco and Peck, 2018; Ranscombe et al., 2019; Kilteni et al., 2012). However, with the multitude of different evaluation metrics that are readily available, a prevailing

limitation amongst related work is the absence of a standardised performance metric, which renders inter-study comparability difficult (Triantafyllidis and Li, 2021a). In the context of understanding the sensory cues that govern motor actions and their effectiveness, one could for instance use a plethora of tools ranging from spatial to time-based metrics. The challenge that encompasses this approach, which unfortunately is prevalent amongst related works, is the lack of consistency according to Triantafyllidis and Li (2021a). This is further aggravated by the complex spatial variables that are entailed when attempting to measure human motor actions, which operate in higher-dimensional space with varying Degree of Freedom (DoF)s (Savage-Knepshield et al., 2016). Perhaps, by fusing two or more metrics under a single formulation in the context of assessing motor performance, a generalised formulation can be derived; not only aiding subsequent inter-study comparability but also aiding in understanding the implications of emerging methods and tools.

Thirdly, even gaining an understanding of how the sensory-motor system of humans can be harnessed and measured, the next question arises and its implications for embodied intelligence. *How can this be utilised in the realm of embodied intelligence?* Even with current advances in embodied intelligence, solving complex sequential manipulation tasks, over a long-horizon, moreover aggravated by the absence of readily available feedback from the environment and attempting to mimic the expertise to which humans are able to achieve such interactions, is currently far from trivial (Davchev et al., 2022; Fox et al., 2019). As long-horizon sequential end-goals are contingent upon a plethora of different sub-tasks, a particular sequential order of actions is necessitated which requires inferring and orchestrating meaningful scene understandings (Triantafyllidis et al., 2023b). In essence, from a biological standpoint, successful sequential manipulation is accomplished when (i) high-level skills are satisfied, (ii) the sensory events are predicted, (iii) the end goals of the task are known and (iv) the sequences of different skills are conceptualised by the brain and more broadly by the nervous system (Flanagan et al., 2006; Ortenzi et al., 2019). Naturally, due to these advantages, biologically inspired embodied intelligence is an ongoing field of research and advances in minimising the gap between humans and machines are perhaps the first building block. Perhaps, the central aim in minimising this gap is by exploiting the capabilities exhibited by humans and allowing intelligence to grow and explore beyond what is readily given to them. As humans possess priory knowledge when learning motor tasks, perhaps intelligent agents can mimic such behaviour in the form of imitation instead of starting from

scratch as suggested by Goldberg (2019). Meanwhile, sole and naive dependence to learn new skills is also misguided (Zaadnoordijk et al., 2022; Zador, 2019; Saxe et al., 2021), due to violating a fundamental neuro-biological principle (Zador, 2019; Saxe et al., 2021), whereby even human infants in their early development, must learn useful representations from unlabeled data and are encouraged to explore in a trial and error manner with their surroundings (Zaadnoordijk et al., 2022). Perhaps, a "middle-ground" is necessitated whereby computational systems can be exploited by the priory knowledge humans possess, yet would still learn and evolve in an unsupervised manner to draw inspiration from a neuro-biological standpoint.

Fourthly, imagining that an understanding of the human sensory-motor system can be established and embodied intelligence rooted in biology developed, the final frontier is the further strengthening of such AI. *The last question that surrounds this aim is: How can the bridge between biology and AI be further solidified and how can embodied intelligence be further improved?* The last dimension that remains relatively unexplored is the power of language and its interplay with cognition. Human cognition is notably enriched by the use of language, especially surrounding the thought process which is capable of eliciting meaningful reasoning to address intricate tasks as supported by Lupyan and Clark (2015). Perhaps, if embodied intelligence is further to be rooted in biology and mimic the extraordinary capabilities exhibited by humans, a mechanism that can harness the instructive and explanatory power of language may be necessary. A two-fold challenge is presented to address this problem. First, understanding how language-based instructions or guidance can be formulated to guide embodied intelligence and most notably robots for intricate tasks spanning over long horizons is necessitated. Secondly, examining how such a language-guided framework can be integrated into already biologically inspired embodied intelligence is needed, to allow further biomimicry by eliciting reasoning with linguistic cues. A promising step towards this vision is the utilisation of Large Language Model (LLM)s with their context and common-sense aware reasoning. LLMs can act as a valuable source for inferring higher-level scene understandings as to provide meaningful guidance in intricate tasks (Du et al., 2023; Yu et al., 2023; Wei et al., 2023). This would not only foster an exploratory edge and bolster the adaptability and capability of robotic systems but also bring them a step closer to bridging the gap between biology and AI.

1.2 Research Aims and Focus

Overall, from the preceding open research problems, this doctoral thesis embarks on a multi-disciplinary exploration, presenting advancements in the fields of (i) Human Factors, (ii) Machine Learning and (iii) Robotics.

Central Research Aims – The preceding advancements are accomplished with a grand total of **four contributions**. These encompass multi-disciplinary research, (i) commencing with the intricacies of the human sensory-motor system and how it can be harnessed, (ii) its motor actions quantified, (iii) to inspire advancements in biologically derived embodied intelligence and (iv) further solidifying AI by integrating language-based insights.

Firstly, with the hopes of harnessing the capabilities exhibited by humans to derive biologically embodied intelligence, understanding the underlying human sensory-motor system is of necessity. Human motor movement is inherently dependent on the sensory system and in particular on multimodality as supported by Billard and Kragic (2019) In fact, to initiate meaningful motor actions, the simultaneous presence of numerous senses is necessitated (A Ghazanfar and E Schroeder, 2006; James et al., 2002; Berger et al., 2018). Unfortunately, to date, a comprehensive overview of how these senses truly contribute towards effective motor actions in the broader context of motor tasks, is largely unexplored (Triantafyllidis et al., 2020). While the determination of sensory cues is largely task-dependent, within the context of manipulation, visual, auditory and somatosensory cues are the most prevalent and most developed (Popescu et al., 2002). In fact, vision, audition and somatosensory stimulation are argued to contribute to approximately 70%, 20%, and 4% towards the human sensory system respectively as suggested by Heilig (1992). However, an extensive comparison of these modalities and their full factorial comparison on motor performance, to the best of the author's knowledge is missing amongst the related work according to Sigrist et al. (2013). Naturally, as part of the first contribution of this thesis and paving the way for understanding how the sensory-motor system works; a comprehensive overview of multimodal interfaces and their contribution to motor task performance is made. To this end, vision, audition and somatosensory feedback states and their combinations are evaluated over increasingly more complex spatial motor tasks. In this way, a thorough understanding of multimodal interfaces entailing visual, auditory and somatosensory states can be made and their correlation in generating effective motor actions across varying complexity tasks established.

Secondly, while harnessing the human sensory-motor system is the first building block, it is of equal importance to consider how meaningful motor actions via appropriate elucidation of the human senses can be adequately measured and quantified. In the broader context of assessing new methods and more specifically evaluating human motor performance to establish correlations and patterns, a plethora of different evaluation metrics can be utilised. However, with the multitude of different evaluation tools that are readily available and the absence of a standardised evaluation tool that can encompass multiple metrics under a single model, inter-study comparability is rendered notably difficult (Steinfeld et al., 2006). This is a prevalent limitation amongst related work attempting to measure and quantify motor movement, primarily attributed to the lack of consistency and the complexities of the spatial variables entailed in higher dimensional space with varying DoFs (Triantafyllidis and Li, 2021a). To this end and as part of the second contribution of this thesis, the author addresses this gap in the literature. A focus on Paul Fitts' original formulation – short for Fitts' law – is given; due to its notable popularity amongst related work and its ability to measure human movement time based on scene layouts; encapsulating in essence time and spatial-based measurements under a single model (Fitts, 1954; Fitts and Peterson, 1964; MacKenzie, 1992). Even though the law's applicability is tailored for lower dimensional movements, the law presents itself as a promising basis for extension. Naturally, due to the popularity of the model; multiple extensions have arisen amongst the related work, which unfortunately do not represent higher dimensional space accurately. Collectively, current work on extending Fitts' law in full 3D space is missing. As a result, a set of increasingly more spatially complex motor tasks are devised to comprehensively evaluate whether the original law, including its extensions, can capture the complex higher-dimensional motor actions exhibited by humans (Triantafyllidis and Li, 2021a; Triantafyllidis et al., 2021; Kumar and Todorov, 2015). To this end, by thoroughly evaluating all spatial variables entailed in such dimensions in contrast to current work, a novel metric is derived that extends Fitts' law to full 3D space entailing 6DoFs movement. As a result, this metric can be utilised to measure and quantify human motor actions in higher dimensional space to better understand the human sensory-motor system and ultimately aid subsequent inter-study comparability amongst related studies.

Thirdly, once the human sensory-motor system is understood in the context of sensory cues translate to effective motor actions which can moreover be quantified and measured, the next aim is to develop embodied intelligence inspired by these biological insights. From

a biological standpoint, humans are proficient in meeting daily environmental demands; especially those concerned with motor tasks (Ashe et al., 2006; Ortenzi et al., 2019). Navigating to a kitchen area and filling a glass with water may seem trivial to humans by virtue of their advanced sensory-motor system, however, from an embodied intelligence this comes with significant challenges. Most notably, current AI systems struggle during intricate sequential manipulation tasks entailing long time horizons and whereby feedback from the environment may not be readily apparent, necessitating notably increased exploration (Davchev et al., 2022; Fox et al., 2019). Moreover, during long-horizon sequential tasks, most prominently seen in robotic manipulation (Davchev et al., 2022; Billard and Kragic, 2019; Triantafyllidis et al., 2023b), the successful completion of an end-goal is contingent upon the successful operation of other sub-tasks; consider a scenario whereby the glass is not readily visible, one must initially retrieve it before drinking from it. To address these challenges, the third and larger contribution of this thesis tackles the challenges of long-horizon sequential tasks. To achieve this, this thesis delves into the domain of hierarchical learning and Mixture of Experts (MoE) architecture. Inspired by the previous two contributions, a biologically-inspired Hybrid-Hierarchical Learning (HHL) architecture is derived – the RObotic MAnipulation Network (ROMAN). ROMAN is a hierarchical learning architecture composed of an ensemble of numerous fundamental and high-level specialising experts, centrally governed and orchestrated as macro-actions by a gating network, referred to as the Manipulation Network (MN). By virtue of ROMAN's hybrid learning and the employed high-level task decomposition, the architecture is tailored for notably intricate robotic manipulation tasks that entail complex settings, sequences and long-horizons. By exploiting the priory knowledge humans possess when learning new motor tasks as suggested by Goldberg (2019), ROMAN blends a plethora of learning paradigms that entail supervised, imitation and reinforcement learning. Internally, the hybrid learning procedure of ROMAN blends in a hybrid manner the exploitation from humans in the form of imitation skills to provide critical decision-making skills and combines the exploratory potential of Reinforcement Learning (RL). The employed hybrid learning is inspired by the previous two contributions and also stands firm from a neuro-biological standpoint whereby task mimicry is encouraged, yet useful representations and learning should still stem from exploration and unlabeled data (Zador, 2019; Saxe et al., 2021; Zaadnoordijk et al., 2022). By comprehensively evaluating the architecture, results will ultimately show that ROMAN firmly retains robustness against increasing uncertainty levels, increased horizons and most importantly exhibits behaviour beyond imitation by

achieving task adaption and autonomous failure recovery capabilities when necessitated. ROMAN's value for intricate of long-horizon tasks requiring adaptive motor skills will be underlined and represents a first step towards minimising the gap between humans and machines, inspiring the derivation of embodied intelligence rooted in biology.

Fourthly, while the gap between humans and machines may be minimised by studying the human sensory-motor system and subsequently deriving biologically inspired embodied intelligence, this gap needs to be further strengthened and solidified. As part of the last challenge and research focus of this thesis, further strengthening and improving embodied intelligence is necessitated. A current challenge that AI faces, whether inspired by biology or not, is the necessity for exploration and the challenges that arise with sparse rewards and intricate long-horizon tasks. While ROMAN's architecture is a promising step forward, additional inspiration from a neuro-biological standpoint may further enhance embodied intelligence. Drawing again inspiration from biology, another important factor to be taken into consideration is the power of reasoning – language. Humans are capable of eliciting reasoning via language by guiding their thought process to address intricate tasks according to Lupyan and Clark (2015). Consequently, by incorporating language in AI perhaps the gap between biology and machines can be further minimised with the hopes of further improving machine intelligence to address intricate settings. The central aim is two-fold, (i) understanding how language-based instructions can guide intelligent agents and (ii) how these can be integrated within already biologically inspired embodied intelligence methods (most notably ROMAN's architecture). LLMs provide a promising step towards this goal, due to their valuable language-based instructions entailing context and common-sense aware reasoning (Du et al., 2023; Yu et al., 2023; Radford et al., 2019). By eliciting reasoning from LLMs (Wei et al., 2023), language-based instructions can provide a clearer and more direct exploratory edge to agents faced with intricate settings. However, LLMs are susceptible to occasional prompt inaccuracies and due to this shortcoming, their direct integration in the decision-making process constitutes language models problematic (Carta et al., 2023; Brown et al., 2020). Perhaps, including LLMs in the existing exploratory paradigm seen with RL, a promising basis for guiding exploration can be devised. To this end, the final part of this thesis delves deep into integrating LLMs with embodied systems, exploring the synergies and potential advancements it can bring to the field of robotics. To this end and as part of the final contribution, a novel language-based framework is presented – the Intrinsically Guided Exploration from Large Language Models (IGE-LLMs). IGE-LLMs is a language-based framework that harnesses the power

of LLMs as an assistive intrinsic reward source alongside the convectional RL paradigm. In essence, IGE-LLMs exploits the reasoning of LLMs for a mere assistive reward, while the exploratory nature of RL still eventually becomes the main policy driver. In this process, the power of language reasoning can be harnessed, to guide the exploration process during intricate settings. Results will ultimately show that IGE-LLMs exhibits notably higher performance over existing related methods, overcomes the shortcomings of direct dependence on LLMs, can be integrated easily and complement other learning methods and is capable of extending ROMAN highlighting its modularity on existing learning architectures. IGE-LLMs is the final contribution of this thesis inspired from a biological standpoint, further narrowing the gap between humans and machines and harnessing the power of language to address intricate challenges in embodied intelligence.

Research Outcomes – Ultimately, the expected outcomes of this research project are with the hopes that the aforementioned advancements provide novel opportunities for innovation and growth across a wide range of fields. These include robotics, embodied intelligence, user experience, human ergonomics and machine learning. Nevertheless, additional research is necessitated and in Chapter 7, the limitations and future work of this research thesis are outlined.

1.3 Research Contributions

The overall contributions of this thesis are summarised below. Contributions I to IV correspond to Chapter 3 to Chapter 6 respectively, while Contribution V is the accumulation of the Literature Review conducted in Chapter 2.

Significant Contribution I: Enhanced Understanding and Harnessing of the Human Sensory-Motor System

- **1.1** Comprehensive full factorial comparison of a plethora of multimodal interfaces entailing visual, auditory and somatosensory states and their effect on motor actions and more broadly on the human sensory-motor system;
- **1.2** Thorough analysis of the effects of uni-modal, bi-modal and tri-modal interfaces on generating meaningful motor actions across varying spatial complexity motor tasks;
- **1.3** Findings establishing a correlation of which sensory modalities contribute to higher usability and decreased, in turn, workload demand in intricate manipulation tasks;

1.4 A reproducible and modular apparatus entailing the integration of a plethora of sensory modalities in a high-fidelity physics simulation closely resembling real-life motor tasks for a comprehensive assessment of the human sensory-motor system and subsequent emerging technologies on human factors and ergonomics.

Significant Contribution II: Leap Towards Standardising and Quantifying Human Motor Actions and Movements

2.1 The establishment of a novel evaluation metric based on a well-rooted in human ergonomics formulation – Fitts' law, that can capture human motor movements and actions in higher-dimensional space of varying DoFs notably better than existing metrics and extensions, quantifying complex bio-mechanical motor movements;

2.2 Thorough comparison of the proposed metric with others amongst the related literature, statistically analysing important spatial variables under increasingly more complex motor experiment tasks, to ultimately evaluate the capability of existing formulations on capturing higher-dimensional human motor actions;

2.3 The proposal of guidelines and recommendations when designing tasks that necessitate careful consideration of human ergonomics, with a notable focus on robotic teleoperation, general motor interactions and 3D user interfaces, particularly when employing MR technologies.

Significant Contribution III: Biologically-Inspired Embodied Intelligence for Hierarchical Task Learning in Intricate Long-Horizon Tasks

3.1 The proposal of ROMAN, an event-based HHL framework that leverages a multitude of different learning paradigms to bring embodied intelligence closer towards the capabilities exhibited by humans from a biological standpoint, as to solve intricate long-horizon sequential tasks in robotic manipulation;

3.2 ROMAN's gating network (the MN) and its ability to achieve a long-horizon sequential end goal by (i) orchestrating robustly a plethora of different macro actions, (ii) exhibiting higher-level scene understandings even in increasingly exteroceptive uncertainties and (iii) exhibiting adaptation and recovery capabilities to beyond demonstrated behaviour;

3.3 By virtue of the employed hybrid learning and higher-level task decomposition, ROMAN's specialising experts are capable of individually surpassing the performance

of whole equivalent hierarchies amongst related work;

3.4 Comprehensive quantitative evaluation of ROMAN's framework against (i) increasing levels of exteroceptive uncertainty, (ii) increased horizons, (iii) number of employed dataset demonstrations, (iv) comparison against monolithic networks and (v) an extensive ablation of the internal learning paradigms.

Significant Contribution IV: Language-Guided Embodied Intelligence for Enhanced Exploration during Intricate Tasks

4.1 The proposal of IGE-LLMs, a framework utilising LLMs as an assistance intrinsic reward source alongside the conventional RL setting, fostering and promoting exploration in intricate environments challenged by sparse rewards and long horizons;

4.2 IGE-LLMs ability to devise and promote a more efficient and direct exploration strategy than existing related intrinsic learning methods and foster computational efficiency during recurring state-action pairs;

4.3 IGE-LLMs overcomes the natural shortcomings of LLMs which are subject to occasional prompt inaccuracies, by leveraging language as mere guidance alongside the exploratory nature of RL, rather than integrating LLMs as direct policy drivers;

4.4 The highlighted modularity of IGE-LLMs due to its ability to (i) be combined and complement existing learning methods and (ii) extend ROMAN's architecture to promote exploration in intricate robotic tasks and automate the orchestration of a plethora of macro-actions with the power of language-guided insights;

4.5 IGE-LLMs's insensitivity to most logarithmic spaced intrinsic scaling parameters equaling and exceeding the extrinsic term, while also exhibiting robustness to increased levels of uncertainty and horizons;

Significant Contribution V: Extensive Multi-Disciplinary Literature Review

5.1 Comprehensive multi-disciplinary review of prior work at the intersection of the human sensory-motor system, machine learning, and robotics. This laid the groundwork for understanding, harnessing and quantifying human sensory-driven motor actions, leading to the development of a biologically-inspired embodied intelligence architecture enhanced with language-guided insights designed to tackle intricate robotic manipulation tasks.

1.4 Thesis Outline

This PhD thesis is organised into a grand total of seven chapters, which are outlined below. Moreover, a visual schematic of the thesis structure is shown in Figure 1.2.

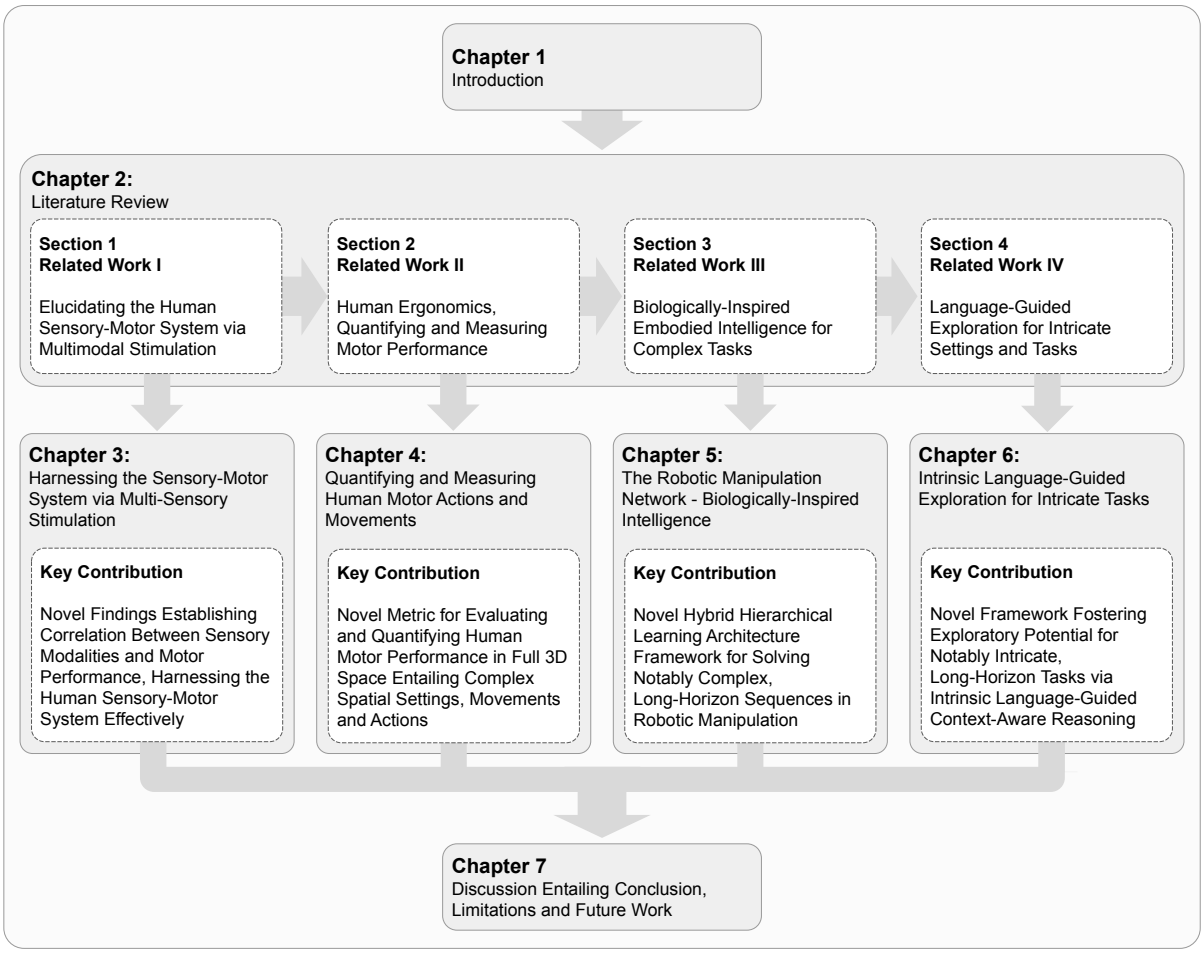

Figure 1.2: **Visual overview of the structure of the thesis.** The figure serves as a visual aid for depicting the structure of the thesis, how the individual chapters are structured and how the sections are inter-connected.

Chapter 1 ***Introduction.*** This chapter. The introduction to this PhD thesis, the problem statement and the motivation of the underlying research aims and objectives.

Chapter 2 ***Literature Review.*** In this chapter, the literature review of the state-of-the-art is presented that is related to this thesis. More specifically, in Section 2.1, the related work on the sensory-motor system of humans is presented, focusing on harnessing this system, coordinating motor actions, and their correlation to sensory stimulation,

with a particular emphasis on multisensory research. Section 2.2 delves into the related work on measuring intricate human motor actions, emphasising the quantification of such movements and broader human ergonomics research. Section 2.3 introduces the state-of-the-art machine learning approaches in the context of embodied intelligence, with a focus on biologically-inspired AI and hierarchical learning for complex tasks and environments. Finally, Section 2.4 presents a literature review on harnessing contextual insights from LLMs, highlighting their usage alongside existing learning paradigms to guide agents in tasks and environments demanding extensive exploration.

Chapter 3: ***Harnessing the Sensory-Motor System via Multi-Sensory Stimulation.*** This chapter outlines part of the initial aim of this thesis and more specifically how different sensory combinations affect motor performance during manipulation tasks. In particular, different combinations of visual, auditory and somatosensory are studied and their corresponding effect on numerously employed objective and subjective human measurement metrics. In Section 3.1 the motivation for studying different combinations of multimodal interfaces is outlined which thereafter formed the hypotheses of the work presented in Section 3.2. Subsequently, in Section 3.3 the methodology is presented in detail and in particular the study design and apparatus, that later allowed for the conceptualisation of the experiments, outlined in detail in Section 3.4. In Section 3.5 the results are presented and in Section 3.6 the inferred findings are discussed with respect to the state-of-the-art, including a brief outline of the limitations and future work.

Chapter 4: ***Quantifying and Measuring Human Motor Actions and Movements.*** Extending on the work of Chapter 3, this chapter presents effective ways of measuring and quantifying human motor performance, using well-established metrics in ergonomics and Human-Computer Interaction (HCI) research. During the work undertaken in the previous chapter, multiple metrics were employed. However, the absence of a standardised metric to effectively capture and most importantly quantify human motor performance was missing. Consequently, Section 4.1 introduces the motivation for deriving a standardised human performance metric that can capture complex motor movements and presents existing metrics that could be used as a basis. Thereafter, the methodology in Section 4.2 is presented and how the study was conceptualised to allow the subsequent evaluation and derivation of the experiments which is described in Section 4.3. Next, Section 4.4 presents the results from the experiments, comprehensively

outlining the different spatial variables entailed in 3D space and how these quantitatively influence motor performance. Most importantly Section 4.5, presents the derivation of a novel metric based on a thorough evaluation of the results, showing improved modelling and quantification of human motor performance in 3D space, which is lastly discussed and framed with the related work in Section 4.6.

Chapter 5: ***The Robotic Manipulation Network – Biologically-Inspired Intelligence.*** While from a biological perspective, solving a complex sequential task in manipulation seems to be straightforward and quickly conceptualised by the nervous system in humans, from an embodied intelligence perspective, this remains a notable challenge. Motivated by the highly advanced human sensory-motor system, comprehensively studied in Chapter 3 and effectively measured Chapter 4, the next aim of this research project was to investigate effective techniques and methods to align the highly advanced perception and reasoning of humans closer to embodied intelligence – robots; in solving complex, long-horizon sequential tasks in manipulation. Inspired from a biological perspective, this chapter introduces ROMAN, a novel hierarchical architecture for solving complex, long horizon sequences in robotic manipulation via hybrid hierarchical learning. Foremost, the motivation and introduction to the problem of solving complex long horizon sequences are presented (Section 5.1). Thereafter the methodology is outlined in (Section 5.2), containing a thorough technical overview of ROMAN, its learning paradigms and characteristics. In Section 5.3 the results of the architecture are presented and subsequently in Section 5.4 these results are inferred and framed with respect to existing findings in the state-of-the-art as well as the identification of the future work and limitations of ROMAN.

Chapter 6: ***Intrinsic Language-Guided Exploration for Intricate Tasks.*** In the final chapter, this thesis explores the usage of LLMs and their potential to provide context and common-sense aware reasoning to guide embodied intelligence in intricate settings, complementing existing human and biological-inspired frameworks. By eliciting the reasoning from LLMs, a novel framework is presented – IGE-LLMs. By leveraging LLMs as an assistive intrinsic reward source, IGE-LLMs guides the exploratory process alongside the conventional RL setting to address complex, of long-horizon and sparse reward robotic manipulation tasks. By utilising and extending ROMAN's architecture (detailed in Chapter 5), it is shown that IGE-LLMs is a modular framework capable of complementing existing hierarchical architectures and learning-based intrinsic methods.

First, a brief introduction on the capabilities and challenges of LLMs and existing intrinsic methods is presented (Section 6.1), rendering the basis of the methodology entailing the derivation and the technical elaboration of IGE-LLMs (Section 6.2). Subsequently, the evaluation of IGE-LLMs and its implementation on two main environments is outlined (Section 6.3), which are then used as the main validation environments presenting the experimental results (Section 6.4). Finally, a brief discussion is presented; tailored to the IGE-LLMs framework and more broadly to the usage of language-guided exploration in intricate settings (Section 6.5).

Chapter 7: ***Conclusion.*** In the final chapter, the main conclusions of this thesis are drawn. In particular, a summary of each individual chapter is provided in Section 7.1, separated into distinct paragraphs corresponding to a chapter. With the same structure and outline, Section 7.2 presents the overall limitations as well as future work for extending the research of this thesis, with the goal of aligning human capabilities closer to embodied intelligence and more broadly, deriving AI from biological insights.

Appendices The appendices include additional materials that were not deemed critical to the main understanding and contents of the thesis. Instead, these appendices are provided to complement the existing contents in the form of supplementary materials. These provide additional technical elaborations and inferring the results in more detail.

Chapter 2

Literature Review

In this section the related work as part of the PhD thesis is outlined and a comprehensive multi-disciplinary review is presented covering three main areas of research (i) Robotics, (ii) Human Factors/Ergonomics and (iii) Machine Learning. A total of four distinct sections of the state-of-the-art is carried out to identify the gaps in current work. These sections form the backbone of the literature review in this thesis and subsequently led to the identification of four open research problems and primary gaps in the existing work.

Section 2.1 introduces a comprehensive review of the human sensory-motor system and how it can be understood and harnessed. In particular, emphasis is given to how humans perceive their surrounding environment and which sensory modalities come into play to initiate effective motor actions to meet environmental and task-specific demands. The review is focused on the broader human sensory-motor system, with a focus on audio-visuo-haptic interfaces and their role in motor performance entailing varying spatial complexity motor tasks. The literature is grouped into the different sensory interfaces that are used in the related work and with respect to the human sensory-motor system. This section forms the literature review backbone for Chapter 3.

Section 2.2 builds on Section 2.1 and outlines the state-of-the-art concerning effective ways of measuring the human sensory-motor system and in particular different existing human performance metrics are reviewed. A notable focus is given towards Fitts' law and

Some parts of this chapter are the accumulation of the author's published works, where the author has contributed as the first author. For a comprehensive list and details of these publications, please refer to the Publication List.

its extensions for quantifying the complex motor actions commonly seen by humans in higher-dimensional space. The literature is grouped based on the different metrics used, their popularity and their distinct applicability towards different spatial measurements. This section forms the literature review for Chapter 4.

Section 2.3 presents a literature review on existing machine learning algorithms and in particular related work with a focus on solving notably long-horizon and intricate sequential tasks as commonly seen in robotic and general manipulation tasks. An emphasis on biologically-inspired embodied intelligence is given based on the preceding segments (Section 2.1 and Section 2.2) identifying the proficiency of humans in meeting intricate environmental demands. The review is grouped based on the different learning paradigms, neuro-biological findings and their implications on embodied intelligence as well as hierarchical learning architectures. This section forms the literature review for Chapter 5.

Section 2.4 delves into a literature review on harnessing the contextual insights and common-sense aware reasoning from LLMs. Emphasis is given as to how LLMs can be utilised to promote exploration and guide agents in intricate settings challenged by sparse rewards and long-horizons. This literature review is more closely framed to the previous section (Section 2.3) with an emphasis on improving learning paradigms especially those that are biologically inspired – fusing the capabilities of human insights and the computational power of AI with language guidance. The literature is grouped from the challenges embodied intelligence faces, especially when employing RL, and progressively reviews appropriate strategies and language-based models and methods in overcoming such shortcomings. This section forms the literature review for Chapter 6.

2.1 Related Work I: Harnessing and Elucidating the Human Sensory-Motor System

With the recent spike in Virtual Reality (VR) and more broadly MR technologies, elucidating the human sensory-motor system is becoming increasingly more immersive. Related work has shown that multi-sensory stimulation in particular, within Virtual Environment (VE)s, can lead to a promising increase in functional motor task performance in simulated remote robotic piloting and manipulation task, according to Triantafyllidis et al. (2020). Nevertheless, during the specific context of remotely inhabiting a robotic system and more broadly an avatar in VEs, the overall experience can to a great extent

feel unnatural. This is due to physical detachment between the operator and the device or body, with the underlying reason being insufficient sensory input and stimulation (Sigrist et al., 2013; Chen et al., 2007).

Humans utilise a wide range of stimuli when perceiving their surrounding environment (Turk, 2014; Quek et al., 2002). In particular, the simultaneous presence of multiple modalities appears to be deeply rooted in the sensory-motor system of humans to generate meaningful motor actions (A Ghazanfar and E Schroeder, 2006; James et al., 2002). In the context of motor movement tasks, visual, auditory and somatosensory feedback cues appear to be the most prominent according to Heilig (1992); Billard and Kragic (2019). In the related work, it is indeed widely considered that manipulation as well as the general ability of humans to meet intricate environmental demands, especially sequential tasks, necessitates multimodality as supported by Bunt et al. (1998); Billard and Kragic (2019). Thus, it would seem natural to steer focus towards ways of effectively maximising the human senses, to illuminate and establish a correlation with how these sensory states affect the generation of useful motor actions. Research in multimodal interfaces and multisensory integration aims to provide effective solutions to increase the feeling of body ownership, namely embodiment and more broadly harness the capabilities exhibited by humans with their intricate sensory-motor system (Yanco and Drury, 2004; DeJong et al., 2004; Jennett et al., 2008). Consequently, understanding the underlying biological factors that govern the sensory-motor system could help illuminate ways of harnessing these capabilities for increased immersion with the hopes of bringing these closer to embodied intelligence.

2.1.1 Multimodal Interfaces and the Sensory-Motor System

Multimodal interfaces can mitigate to a great extent the high complexities that inherently surround meeting environmental demands and interactions as evidenced by Triantafyllidis et al. (2020). In particular, the work of Triantafyllidis et al. (2020) has shown that the additions of multiple sensory feedback states, such as bi-modal or even more so tri-modal feedback correspond to notably higher motor performance even when necessitating the fulfilment of increasing intricate spatial demands. The value of multimodal interfaces has also been shown during overloaded sensory cues. In particular, during visual information overload, offloading visual stimulation via alternative means of sensory feedback, such as auditory stimulation, can reduce visual cortex stress, corresponding to less workload

demand and ultimately to improved task performance (Lathan and Tracey, 2002; Burdea et al., 1996). These findings are also in line with Billard and Kragic (2019), supporting that manipulation by itself is multi-modal and the simultaneous presence of multiple modalities contributes to the generation of meaningful motor actions.

While the utilisation of specific sensory cues is largely task-dependent, visual, auditory and somatosensory cues are the most prevalent within the context of manipulation and are overall most developed according to Popescu et al. (2002). Less "prominent" sensory types such as olfaction or even gustation have shown promising potential in robotics, however, still remain to a large extent unexplored (Keshavarz et al., 2015; Narumi et al., 2011). According to Heilig (1992), the primary reasons these two types of sensory states remain unexplored are (i) their limited relevance when concerned with motor tasks, (ii) the complexities that surround the design of an apparatus accommodating these and (iii) being less overall developed and contributing less to the overall human perceptive system compared to audio-visuo-haptic stimuli. In particular, Heilig (1992) estimated that out of the five human senses (excluding proprioception), there are different sensory contributions for each towards the human sensory-motor system. In their work, Heilig (1992) quantified the contribution of these to be roughly 70%, 20%, 5%, 4% and finally 1% for sight, hearing, smell, touch and taste respectively. Consequently, the review of the state-of-the-art in this thesis will be narrowed down to the sensory modalities of vision, audition and somatosensory. Nevertheless, even establishing a correlation between these modalities and their effect on motor performance, it is worthwhile to point out that human perception is a complex phenomenon. In particular, a plethora of secondary influencing factors come into play, which could further influence motor performance. Some of these include the individual's previous exposure to technologies, intricate personality-related factors and more (McGlynn and Rogers, 2017; Barfield, 1995; Sacau et al., 2008; Menchaca-Brandan et al., 2007). While accounting for all of these may prove to be notably challenging, interdisciplinary efforts stemming from psychology may aid in unravelling some of these intricacies.

In terms of more readily available applications, multimodal interfaces have proven their advantages in motor learning and in particular in human ergonomics and robotic teleoperation research as seen in the works of Sigrist et al. (2013); Rosen et al. (2019). During the latter research area, a human operator in essence embodies the robotic system. In this process, the human and the robot are physically detached from one another, which can, in

turn, constitute initiated motor actions and more broadly the "embodiment experience" unnatural (Triantafyllidis and Li, 2021b). Without effective methods of increasing one's feelings of body ownership via increasing perception and situational awareness (amongst others), motor task performance during such cases has shown to be limited according to Triantafyllidis et al. (2020). Consequently, multimodal interfaces appear to be an appropriate mitigation strategy for this problem. Research has shown that carefully designed sensory feedback states that if well synchronised with one another, can not only complement each other as supplementary stimuli-induced-information, but most importantly contribute towards notably increased motor task performance, especially in emote robotic piloting (Chen et al., 2007; Yanco and Drury, 2004; DeJong et al., 2004; Jennett et al., 2008).

Visual Stimulation

Dominance and Importance of Vision – Vision is the predominant sense in humans, as (i) information can be extracted substantially faster from the environment than the other senses and (ii) it dominates the other senses when sensory conflict arises (Heilig, 1992; McIntire et al., 2014; Klatzky et al., 1998). Heilig (1992) even quantified that vision alone accounts for about 70% towards the sensory contribution in humans. Rock and Victor (1964) further observed that when sensory conflicts between somatosensory and visual stimuli arise, vision is strongly dominant and outweighs somatosensory feedback. Moreover, the findings of Klatzky et al. (1998) also showed that visual feedback can dominate vestibular cues, similarly to somatosensory. The implications of Klatzky et al. (1998) are rather of notable importance for motor actions and generally manipulation tasks, since the vestibular system is responsible for the balance and spatial orientation of one's system, deeply rooted in coordinating movements with balance. To further highlight the importance of carefully designing visual feedback and the implications on other senses, Burns et al. (2007) found that vision is furthermore dominating proprioception. Proprioception, also referred to as kinaesthesia constitutes the "sixth human sense", which is the sense of self-movement and the relative joint position of one's body (LaValle, 2019). Consequently, as it can be inferred, the dominance of vision is well supported by current research. Hence, carefully designing visual interfaces should be of high importance during the presence of multimodal feedback.

Popular Types of Visual Stimulation – The two most popular methods of providing visual stimulation are either through **monocular** or **stereoscopic** vision. The two most widely used types of visual devices are generic display monitors for monocular vision and Virtual Reality Head Mounted Display (VRHMD)s for stereoscopic vision according to Triantafyllidis et al. (2020). Existing work has extensively investigated both types of visual stimulation and there is significant evidence that VRHMDs provide a notable increase in human performance during object interaction and manipulation (Azmandian et al., 2016; Slater et al., 1996; Poupyrev and Ichikawa, 1999). This appears to hold especially true during telepresence and teleoperation control as supported by Tachi (2010); Rae et al. (2015). Nevertheless, an extensive comparison of monocular and stereo vision, alongside other sensory combinations is missing in the related work.

Depth Information and Vision-Based Associated Variables – One important aspect within the umbrella of visual feedback is depth estimation and Field of View (FoV). Numerous studies have provided evidence that distance estimation for physical targets is better than those for virtual targets as seen in simulation environments (Lin and Woldegiorgis, 2015; Durgin et al., 1995; Witmer and Kline, 1998; Swan et al., 2015). In particular, other studies have shown that humans overestimate depth-associated distances in virtual environments (Lampton et al., 1995; Renner et al., 2013). However, the discrepancies surrounding the estimation of depth between real and virtual environments and drawing conclusions from these is still an intricate phenomenon. This is particularly due to perception being inherently difficult to quantify and subject to each person's visual health, mostly correlated to age (Hess et al., 2015; Zaroff et al., 2003). Thankfully, accounting for these human factors is possible during initial participant recruitment and consistency during the selection criteria, which can account for these confounding factors according to Triantafyllidis and Li (2021b). Nonetheless, some studies have attempted to enhance depth perception via simulated approaches, such as rendering checkerboard textures in the simulated environment, however, no notable effects were observed as shown by Deng et al. (2019). VRHMDs have to a great extent provided an appropriate solution to this problem, by offering increased overall depth perception and environmental awareness than standard monocular monitors (Rosenberg, 1993; Triantafyllidis et al., 2020). Increasing overall depth information, in turn, has proven to lead to reduced collisions with the surrounding environment and corresponding to notably better motor performance during dexterous manipulation tasks (Martins and Ventura, 2009; Scribner and Gombash, 1998).

In addition to depth-related information being a significant factor, FoV, also has a notable effect on performance. The related work supports that overall larger FoV can lead to increased performance and environmental awareness (Scribner and Gombash, 1998; Smyth, 2000; Johnson et al., 2015). However, decreased usability and increased mental demand can on the other hand be observed according to Johnson et al. (2015).

From the above, it is inferred that vision is a dominant sense and should carefully be accounted for during the designing of multimodal interfaces. Moreover, it is identified that the two most widely used methods of providing visual stimulation are either through monoscopic or stereoscopic viewing. It is also concluded that depth information and FoV are two important vision-based variables that should be considered when designing visual stimuli, especially during the selection of appropriate devices that would thereafter accommodate these.

Auditory Stimulation

Importance of Auditory Feedback and its Supplementary Role– In most studies concerned with manipulation, auditory feedback does not constitute a replacement for vision, however, it is rather added as an additional supplementary modality. In particular, related studies have investigated, to a great extent, the effects of bi-modal interfaces entailing auditory and visual cues, also known as audiovisual interfaces (Tachi et al., 2003; Nagai et al., 1999; Shilling and Shinn-Cunningham, 2002; Sigrist et al., 2013). Auditory feedback has been shown to provide enhanced temporal, spatial, amplitude, and frequency resolution, while also acting as a promising supplementary sense, especially when visual information is already overloaded according to Sigrist et al. (2013). Moreover, during high visual load, auditory feedback can increase environmental awareness for human operators (Shilling and Shinn-Cunningham, 2002), correlated to increased performance and decreased environmental collisions in robotic teleoperation (Nagai et al., 1999). In particular, one study found that when FoV is limited during visual stimulation, auditory cues can present themselves as a notable benefit. This is due to operators being able to localise the spatial position of sounds even when not directly gazing at them (Simpson et al., 2013). In another work, Tachi et al. (2003) found that during a First Person View (FPV) manipulation of a humanoid robot, teleoperation control was more intuitive when presented with an audiovisual interface.

Types of Auditory Stimulation – According to Sigrist et al. (2013), there are in general, three major categories of auditory feedback that have been shown to enhance motor learning (i) **Auditory Alarms**, (ii) **Sonification of Movement Variables** and (iii) **Sonification of Movement Error**. In a more general context, auditory information can be represented either in a *terminal* or *continuous* manner. A notable example of terminal feedback would be stimulating audio after a desired event takes place, such as satisfying certain spatial criteria. On the other hand, continuous i.e. concurrent feedback would be stimulated during a given task (Sigrist et al., 2013). The findings of Secoli et al. (2010) revealed that instead of representing significant sensory information via a single channel such as vision, auditory stimulation in a terminal form such as alerts or warning sounds, can decrease overall distractions. The works of Erni and Dietz (2001); Wellner et al. (2008) have shown that continuously controlling auditory pitch, may aid during object clearance and signify the benefits of sound sonification.

Overall, it would appear from the related work that the specific form of auditory representation is important. Nonetheless, the presence of auditory feedback as a whole may provide a richer environmental experience and awareness, ultimately constituting sound as a valuable supplement to just relying on vision. Nevertheless, more investigation would be needed to further confirm these findings and in particular underline the value of auditory feedback beyond vision, for example, when coupled with the addition of somatosensory feedback.

Somatosensory Stimulation

Importance of Somatosensory Feedback – Studies have shown that somatosensory, known more broadly as haptic, stimulation is crucial for improving the performance of a remotely-controlled robotic system that is involved in time-sensitive search and rescue missions (Martins and Ventura, 2009). With increased somatosensory input, one study showed that this correlates to higher levels of situational awareness, leading to fewer collisions and ultimately faster completion times in real-life (Martins and Ventura, 2009) and VEs (Brickler et al., 2018). Somatosensory feedback has been utilised in combination with visual interfaces, which are dominant in many contexts, to further enhance the overall motor movement and manipulation experiences. This has led to the development of visuohaptic interfaces, whereby both vision and haptics are integrated, leading to increased overall spatial awareness (Rock and Victor, 1964; Klatzky et al., 1998; Burns et al., 2007; Bolopion and Régnier, 2013). In particular, the fusion of vision and haptics

has been shown to improve human performance in tasks such as nano-robotics (Ferreira and Mavroidis, 2006), micro-manipulation (Salcudean and Yan, 1994; Fukuda et al., 1988), and the remote control of robotic systems (Cholewiak and Collins, 2000; Rochlis and Newman, 2000). Brickler et al. (2018) found that during a pick-and-place task, the presence of somatosensory feedback notably improved performance, resulting in overall faster mean completion time and faster performance compared to when somatosensory feedback was not present. In a further study, electro-tactile feedback was used in combination with VRHMD to control a robotic arm remotely (Pamungkas and Ward, 2013). In the field of planetary and space exploration, a visuohaptic interface was used for remote manipulation by communicating contact between the robot gripper and an object, with the intensity of vibration indicating object proximity, leading to increased spatial accuracy for the user as evidenced by Aleotti et al. (2002). Lastly, tangible user interfaces, which provide direct somatosensory stimulation through real-world objects, have also been well-researched. These interfaces offer a notably more immersive experience by aligning the shapes of virtual and real-world objects into one (Azmandian et al., 2016; Ullmer and Ishii, 2000). However, a drawback of these interfaces is the need to carefully match the sizes of virtual and real-world objects, necessitating a well-designed apparatus.

Types of Somatosensory Feedback and Devices – As it can be inferred, current research has focused notably on the effects and importance of somatosensory feedback due to its benefits as a complementary state. Consequently, due to its importance, various types of somatosensory feedback devices have been developed, some of which include electro-tactile, vibrotactile, and force-feedback devices. The most commonly used type in the related work is vibrotactile stimulation, which is popular due to being cost-effective and to a great extent straightforward to implement (Kim et al., 2017; Triantafyllidis et al., 2020; Aleotti et al., 2002). In one study, the effects of vibrotactile stimulation were explored using a custom-built somatosensory glove, which consisted of a multitude of coin-type vibration motors and resistors to simulate heat from virtual object manipulation (Kim et al., 2017). Moreover, in diagnostic surgery entailing simulators using VR technologies, more advanced and sophisticated force feedback techniques have been developed to reflect the realistic reaction forces of deformable objects such as soft tissue commonly seen in medical and surgical settings (Vuskovic et al., 2000). Moreover, research into kinesthetic force feedback has demonstrated some advantages over lower-cost methods (Sinclair et al., 2019; Gu et al., 2016). Additionally, force-feedback devices are

capable of constraining the grasping motion of a user's hands based on the virtual object shape being held, rendering them useful for haptic exploration and shape determination as supported by Choi et al. (2016). Nevertheless, as counter-intuitive as it may be, increased somatosensory feedback especially in quality, does not guarantee an increase in performance over lower-cost solutions according to Berger et al. (2018). This is analysed separately in this section titled "Uncanny Sensory Valley and Effects of Overlapping Stimuli".

Audio-Visuo-Haptic Interfaces

To date, no extensive full factorial comparison has been conducted as to how audio-visuo-haptic interfaces and the combinations of these pairs affect motor performance to the best of the author's knowledge. With the aim of understanding how such sensory cues can be harnessed and how the sensory-motor system of humans can be utilised to its fullest potential, a first step would be to understand the implication of multi-modal stimulation on generating meaningful motor actions. The work of Sigrist et al. (2013), highlighted the potential of a multi-modal interface in motor learning and how it can potentially affect functional task performance with an increased amount of motor complexity. In their work, Sigrist et al. (2013) hypothesised, however, ***did not confirm***, that a multi-modal interface becomes notably more effective as the task complexity increases in motor learning via tri-modal feedback entailing visual, auditory or/and haptic feedback. In a meta-analysis, the work of Burke et al. (2006) identified that adding more sensory modalities is correlated to improved performance during motor tasks and appears to be more beneficial than relying on a single sensory channel. The work of Lathan and Tracey (2002) studied certain combinations of visual, auditory and somatosensory through vibrotactile feedback in a telerobotic navigation task, yet did not exhaustively study all their combinations. Moreover, the work of Frid et al. (2018) also explored the effectiveness of audio-visuo-haptic interfaces to assess motor performance in visual throwing tasks. However, in both instances of Lathan and Tracey (2002); Frid et al. (2018) a full factorial comparison of all possible combinations of these three modalities was not performed, nor was their effectiveness on increasing motor task complexity studied. Unfortunately, this appears to be important and worth investigating according to Sigrist et al. (2013). To the best of the author's knowledge at the time of writing and the hypotheses formed by Sigrist et al. (2013), tri-modal feedback entailing visual, auditory and somatosensory stimulation and their full factorial comparison on motor performance is still a gap in the literature.

This gap is addressed in Chapter 3, whereby empirical evidence is provided to the untested hypothesis of Sigrist et al. (2013), by exhaustively comparing all possible combinations of vision, audition and somatosensory feedback in manipulation tasks of increasing spatial complexity.

2.1.2 Design Considerations for Multimodal Feedback

When designing multimodal interfaces entailing a plethora of sensory cues, certain important aspects need to be accounted for. In particular, in order to provide the benefits of multisensory feedback as identified to this point, important characteristics need to be considered, which are reviewed and identified from the related work below.

Multimodal Sensory Synchronisation – Foremost, during multimodal feedback, sensory synchronisation is crucial. When sensory signals via different stimuli are out-of-synchronisation, the overall spatial and temporal benefits one would expect from multimodal interfaces are reduced (Popescu et al., 2002). Moreover, **sensory conflict** can arise when different sensory signals in multimodal pipelines are unable to attain synchronised stimulation of all incorporated feedback channels. This, in turn, can lead to decreased motor performance, effectively nullifying the benefits of such interfaces (Richard et al., 1994).

Design and Specific Implementation of Sensory Feedback – As identified to this point, there are a plethora of different design options when designing a particular sensory modality, with terminal and concurrent feedback being the most popular, yet not all, categories (Sigrist et al., 2013). These design decisions appear to be important according to the work of Shang et al. (2019), who also found that concurrent/continuous feedback instead of terminal feedback, can be counterproductive in terms of subjective and objective performance as user preference and time efficiency respectively can be reduced. However, these design decisions may entail numerous confounding variables and may not only be specific to the type of sensory modality incorporated but also to the combinations of other sensory channels, which aggravates generalisation. More research towards this direction may be beneficial in illuminating the underlying effects of designing sensory feedback states and whether these would then be task-specific.

Uncanny Sensory Valley and Effects of Overlapping Stimuli – The term "Uncanny Valley" refers to the break in immersion when an artificial being appears *too* realistic, causing a corresponding negative response from humans towards it (Mori et al., 2012). The work of Berger et al. (2018) explored this notion further in the context of sensory feedback and introduced the notable term "Uncanny Valley of Haptics". In their work, Berger et al. (2018), found that adding somatosensory feedback to a multimodal interface does not necessarily increase task performance and can on the other hand lead to reduced subjective realism. They argued that this phenomenon occurs when somatosensory resolution does not match or coincide with the resolution of other incorporated sensory channels. Neuroimaging studies have provided evidence to this argument, demonstrating that there is a convergence and overlap of visual and somatosensory stimulation in the occipital region of the brain and there may be some inter-dependence between these, (A Ghazanfar and E Schroeder, 2006; Amedi et al., 2001; James et al., 2002; Sathian and Zangaladze, 2002). The implications of these findings are of importance. In particular, these findings imply that even when using high-resolution somatosensory feedback via expensive and complex haptic-capable devices over lower-cost ones, if the resolution is not matched with the other incorporated sensory channels in a multimodal interface, these benefits are nullified. While matching such resolution across all sensory channels is an important finding and should be accounted for, being able to quantify this relationship and ensure that resolution is matched can be notably difficult. Moreover, whether this applies to other sensory channels other than somatosensory stimulation needs further investigation and neuroimaging studies may be necessitated to establish whether there are further overlaps of other modalities in other regions of the brain and their implications.

2.2 Related Work II: Quantifying and Measuring the Human Sensory-Motor System

To this point, in Section 2.1 the importance of maximising motor performance has been reviewed and identified that multimodal interfaces and different sensory feedback states can benefit and harness the human sensory-motor system. By increasing sensory stimulation via multi-sensory integration, overall motor performance can be increased, leading to better situational awareness with direct benefits in tasks necessitating human motor control, with the most notable being remote teleoperation and simulated environments, especially entailing MR technologies (Triantafyllidis et al., 2020).

However, an important problem still remains, can these bio-mechanical motor actions be translated into measurable metrics? Can these be quantified? How can one ensure consistency amongst related work attempting to measure motor actions and movements with a common evaluation metric?

As identified, with the exponential spike in technological growth in both MR technologies and realistic simulators, there has been a notable increase and demand for elucidating the human sensory-motor system. However, a metric capable of quantifying human performance and in particular motor movement in full 3D space is mostly missing and unaccounted for (Triantafyllidis et al., 2020; Drewes, 2010; Barrera Machuca and Stuerzlinger, 2019). Assessing the performance of human motor movements and actions during teleoperation and MR is a challenging problem, particularly in 3D space due to the complex spatial settings entailed. Despite the presence of a multitude of metrics in the related work, a compelling standardised 3D metric is still missing, aggravating inter-study comparability between different studies (Triantafyllidis and Li, 2021b). The absence of such a metric is primarily attributed to the discrepancies between pointing and manipulation, the notably intricate spatial variables associated with 3D space and the combination of translational and rotational movements as seen with higher DoF movements altogether (Triantafyllidis and Li, 2021a).

Fitts' law is perhaps the most widely used human prediction model in HCI history attempting to capture human movement in lower dimensions (Fitts, 1954; Drewes, 2010). Since the formulation's proposal in 1954 and despite the collective effort towards deriving an advanced extension of a 3D human performance model based on it, there is still the absence of a standardised metric that encompasses the plethora of spatial complexities encountered in full 3D space, consult Triantafyllidis and Li (2021a).

In this section, the motivation for deriving a standardised metric that can measure human motor performance is presented. Moreover, the rationale for emphasising Fitts' law is outlined and the investigation of the most prominent extensions of Fitts' law is analysed. Ultimately, the value of extending the simplistic and applicable mostly in lower-dimensional Fitts' law is underlined. Furthermore, the state-of-the-art and most popular model extensions are thereafter compared with their characteristics comprehensively analysed to ultimately allow the derivation of a higher-dimensional metric by accounting for the intricate spatial aspects in 3D space. Lastly, some of the complexities, frontiers and potential challenges that may lie ahead are discussed.

2.2.1 The Simplicity Yet Promising Basis of Fitts' Law

Paul Fitts proposed his original model in 1954, which soon became known as simply **"Fitts' law"** (Fitts, 1954). Fitts' law became and still is to date the most widely used and researched human predictive models in HCI and ergonomics history (Gillan et al., 1992; MacKenzie, 1992; Drewes, 2010; Hoffmann, 1991; Argelaguet and Andujar, 2013; Kerr, 1973; So and Griffin, 2000).

Fitts' original formulation from Fitts (1954); Fitts and Peterson (1964), serves as a human predictive model, predicting the time, also known as *Movement Time (MT)*, based on the spatial layout of a simple 2D scene, known as an *Index of Difficulty (ID)*. The model is formulated as follows:

$$\begin{aligned} MT &= a + b \cdot ID, \\ ID &= log_2\left(\frac{2A}{W}\right) \end{aligned} \tag{2.1}$$

In more simplistic terms, the model predicts the time to reach, point or click to a target location, given the target's distance from the origin of the cursor/hand or object (A), as a ratio of the target's width (W). The logarithmic term ID, represents the index of difficulty measured in bits per second bit/s while the resultant MT is measured in seconds. The constants a and b represent the intercept and slope respectively, both of which are derived via regression.

Importance of the Law The ability of the law to combine both time and spatial-based metrics under a single formulation renders the pursuit of extending it to 3D alone of notable importance (Triantafyllidis and Li, 2021a). The rationale behind this importance stems from the plethora of different evaluation metrics that are readily available in the related work. In particular, relying on a multitude of different types of metrics to assess human motor performance in either MR, VEs or teleoperation, should ideally be avoided, as comparability between the results of such studies is rendered notably challenging (Triantafyllidis and Li, 2021a; O'Hara, 1987). However, if one is able to use an existing formulation such as Fitts' law combining two or more types of metrics under a single model and extending it to higher dimensions, standardisation and comparability of results among related studies will be rendered notably higher. For instance, earlier works on teleoperation as O'Hara (1987) and recent ones from Fani et al. (2018) utilising anthropomorphic robotic hands (19 DoFs) leveraging MR technologies, unfortunately, do

not make use of metrics such as Fitts' law and instead rely on multiple time-/spatial-/behaviour-based metrics. Another important reason why an extension derived from Fitts' law would notably increase inter-study comparability, is due to its popularity and by extent notable focus amongst related work, as well as the formulation itself being the basis of a plethora of subsequent extensions derived (Triantafyllidis and Li, 2021a; Drewes, 2010).

2.2.2 The Problem and Simplicity of Fitts' Law

Nevertheless, despite the formulation being very popular and notably researched, the original formulation by Fitts' as shown in Equation 2.1 is limited in terms of simplicity when full 3D space is considered. In particular, the model is limited to **four-key** areas, namely:

- **Dimensionality** – Considering all possible spatial variables in 3D space. The law is originally derived and applicable for 1D and 2D space.
- **Spatial Arrangements** – Considering spatial arrangements entailing directional and inclination variations around the user. The law and most of its extensions only investigated single-line movements.
- **Combined Movements** – Assessing translational as well rotational types of movements, both of which are fundamental during human motor movement. Most extensions to date have exclusively considered translational tasks as per the original model and investigations in the rotational domain are notably limited.
- **Types of Interactions** – To the best of the author's knowledge, Fitts' human predictive model and all subsequent extensions have exclusively investigated pointing tasks without the addition of physical properties such as gravity or friction entailing physically manipulating objects. Whether it is plausible to extend the law to the manipulation domain remains to be comprehensively tested.

During any type of human motor interaction, most notably entailing MR-based technologies in VEs or general teleoperation, all four aspects are fundamentally inseparable and inherent in human motor movements (Triantafyllidis and Li, 2021a; Kulik et al., 2020; Triantafyllidis et al., 2020). Unfortunately, Fitts's original model, and the subsequent model extensions to date, have still not **collectively combined** all of the aforementioned. In the remainder of this section, the most widely used model extensions based on Fitts' law

are presented. Furthermore, an outline of potential future challenges as well as important research directions is presented, with the hopes of deriving a standardised evaluation tool that can quantify human motor actions and movements.

2.2.3 Extending the Law to 2D and 3D Space

Due to the popularity of the human predictive model formulated by Fitts, numerous extensions started to emerge. In this section, the most popular, and most referenced amongst the related work extensions resulting from the original law are outlined.

Extensions in 2-Dimensional Space

While originally developed for 1D tasks, Fitts' human predictive model has also been widely applied to 2D pointing tasks. Hoffmann (1995) derived an extension based on the original law via a series of discrete tapping tasks using the participants' fingers as pointing probes, with the width of the finger added to the ID of the original law (Equation 2.1) as follows:

$$MT = a + b \cdot \underbrace{log_2\left(\frac{2A}{W+F}\right)}_{ID} \tag{2.2}$$

In this formulation, Hoffmann (1995) the formulation of Fitts' law was retained mostly intact, with the exception of adding the variable F to the equation, representing the index finger pad size of each participant. This formulation opened new and interesting paths and indirectly motivated future work in including the object size, when concerned with manipulating objects, in VEs (Wang and MacKenzie, 1999).

Another popular extension stemmed from the work of Welford (1968), with the most notable change being the omission of the multiplication by two for the target separation i.e. distance A as seen in the original law (Equation 2.1). Instead, the addition of $+0.5$ was presented, such that the formulation was altered as:

$$MT = a + b \cdot \underbrace{log_2\left(\frac{A}{W} + 0.5\right)}_{ID} \tag{2.3}$$

The rationale of Welford (1968) of adding the $+0.5$ term, was to account for the distance from the centre of the target to its edge. The resultant formulation has then been demonstrated to do well in 2D task settings.

Similarly, to Welford (1968), MacKenzie (1992) extended Fitts' original law what soon became to be known as the **"Shannon"** equation and the basis of a plethora of subsequent extensions, especially in later attempts to cover the 3D space. Unlike Welford (1968), MacKenzie (1992) added +1.0 in his formulation, forming their equation as:

$$MT = a + b \cdot \underbrace{log_2\left(\frac{A}{W} + 1\right)}_{ID} \tag{2.4}$$

Overall, this equation yielded an overall better fit and the robustness and linear fit of the Shannon model have been well demonstrated for both translational (Stoelen and Akin, 2010; Murata and Iwase, 2001) as well as rotational tasks (Stoelen and Akin, 2010). The resulting IDs of MacKenzie (1992) are about one bit less than with Fitts's formulation. Nevertheless, contrary to Fitts' original model (Equation 2.1), the Shannon formulation (Equation 2.4) is to some extent limited in terms of mathematical expressiveness. While MacKenzie (1992) argued that the addition of the +1 term avoids negative IDs as with the original Equation 2.1, this appears to not be entirely true. A negative ID in Fitts' model would in theory mean that the cursor or probe is already within the target area. This argumentation limits, to some extent, the theoretical justification for the purpose of higher model fitting as suggested by Drewes (2010). Nonetheless, adopting the Shannon model facilitates the comparability of numerous extensions as this has become the norm for most of the subsequent model extensions presented.

Extensions in 3-Dimensional Space

More recently, Fitts' original model itself has also been tested whether its applicable in the 3D domain (Stoelen and Akin, 2010; Kulik et al., 2020), providing robust grounds for various 3D extensions as seen from the works of Murata and Iwase (2001); Barrera Machuca and Stuerzlinger (2019); Cha and Myung (2013). Even though applicable for lower dimensions, Fitts' original law and subsequent extensions had less, yet still high predictive powers for higher dimensional space. Nevertheless, while Fitts' original formulation has been applied and extended towards the 3D domain by accounting for certain spatial arrangements entailing directions and inclinations, (Cha and Myung, 2013; Murata and Iwase, 2001), current methods to date still do not represent higher dimensions accurately (Triantafyllidis and Li, 2021a; Stoelen and Akin, 2010). This is due to current work not accounting for all the associated spatial variables entailed in the 3D domain, as outlined in Section 2.2.2.

The work of Murata and Iwase (2001), introduced directional angles as the means of covering and accounting for the spatial arrangements as seen when transitioning from 2D to 3D space. In their work, directional angles were introduced under the spherical coordinate system and defined as azimuth angles. Their derived model, based in large part on Equation 2.4, is formulated as:

$$MT = a + b \cdot \underbrace{\left(log_2 \left(\frac{A}{W} + 1 \right) + c \cdot \sin\theta \right)}_{\text{ID}} \tag{2.5}$$

whereby θ represents the azimuth angle, which is effectively added to the ID. The azimuth angle (θ) is furthermore controlled by a constant c, which is determined through regression analysis. Note that their model and contrary to the subsequent extensions presented, the ID not only encompasses the logarithmic term but also the sinusoidal term with the azimuth angle added to it. The under-brace in Equation 2.5 illustrates the ID. In their study, Murata and Iwase (2001) investigated a grand total of eight different levels of directional (θ), or azimuth, angles ranging from 0° up to 315°, with a 45° increment. Interestingly enough, their findings revealed that these angles had a sinusoidal relationship with movement time (MT). In particular, they found that upper (90°) and lower movements (270°), were significantly more difficult, increasing MT notably higher than left (180°) or right movements (0°).

Inspired by these findings, the work of Cha and Myung (2013) conducted a series of experiments by adding inclination angles to their tasks representing additional dimensions in finger-aimed pointing tasks, similar to the spherical coordinate system. Their model extension was based on Hoffmann (1995) as seen in Equation 2.2, formulated as:

$$MT = a + b \cdot \theta_1 + c \cdot \sin\theta_2 + d \cdot \underbrace{log_2 \left(\frac{2A}{W + F} \right)}_{\text{ID}} \tag{2.6}$$

From the above, it can be observed that their formulation encompasses the finger pad size of the participants acting as the pointing probe (F) as seen with Equation 2.2. Moreover, the directional angle ($\sin\theta_2$) is introduced as with Equation 2.5, with the addition of θ_1 representing inclination angles. The constants (a, b, c) and (d) are again determined through regression analysis. From their work, Cha and Myung (2013) foremost confirmed the work of Murata and Iwase (2001) and in particular that directional angles do indeed appear to have a sinusoidal relationship with (MT). Unlike directional angles, they observed that in their study inclination angles shared a linear relationship with (MT),

similar to the distance A. More specifically, upper movements appeared to increase (MT) unlike downward movements. It is furthermore worthwhile to point out that unlike Murata and Iwase (2001), the work of Cha and Myung (2013) limited their investigation to forward motions covering between $\pm 30°$ and $\pm 60°$ azimuth angles.

Nonetheless, while novel findings in regards to directional and inclination angles were presented, neither Murata and Iwase (2001) nor Cha and Myung (2013) investigated an experimental setting entailing both translational and rotational movements, or entailing 3D displays, let alone VRHMDs. This was later accounted for by Barrera Machuca and Stuerzlinger (2019), though limited again for translation tasks only, whereby experiments entailing pointing tasks with the use of 3D displays were conducted with motions ranging from azimuth angles of $-90°$ to $90°$ and $0°$ to $180°$. Their findings foremost confirmed left-to-right movements were easier than movements away from or towards the user as with Murata and Iwase (2001) and interestingly enough, furthermore revealed that depth-associated changes i.e. the distance of the user's eyes to the screen, linearly affected MT when using 3D displays. In particular, they observed that higher depth values were correlated to higher difficulties (ID) and by extent higher timings (MT). Consequently, to account for the depth changes and their significant effect towards MT, Barrera Machuca and Stuerzlinger (2019) formulated the following equation, based primarily on Equation 2.4 as:

$$MT = a + b \cdot \underbrace{log_2\left(\frac{A}{W} + 1\right)}_{\text{ID}} + c \cdot CTD \tag{2.7}$$

In this equation, the term CTD is introduced representing the Change in Target Depth, measured in centimetres. As it can be observed similarly to all aforementioned extensions, this new term is controlled by a constant c, similarly determined through regression. It is worthwhile to point out that the work of Barrera Machuca and Stuerzlinger (2019) limited their investigation to directional angles exclusively, disregarding inclination angles. However, as it was inferred from the related work of Cha and Myung (2013), inclination angles appear to notably affect MT. Consequently, it is important to study the effects of both directions and inclinations in one study encompassing 3D tasks with MR technologies with the hopes of increasing generalisation when deriving new formulations of metrics.

So far in this review, no formulation has investigated **rotational variations** or moreover **combined** rotational with translational movements all in one setting. This is presented in the next section.

Accounting for Combined Movements

To this point, all derived equations either extended the original Fitts' law from 2D to 3D space, yet, were solely limited to translational movements. However, rotation is a fundamental part of the everyday interactions of humans with their environment, as rotation is inherently fundamental in meeting and satisfying certain spatial criteria when meeting environmental demands (O'Hara, 1987; Billard and Kragic, 2019; Triantafyllidis et al., 2020). Combining translational and rotational movements under one setting, and ultimately in one unified formulation, is essential and an inseparable part when manipulating either real or virtual objects in either the 2D or the 3D domain (Kulik et al., 2020; Stoelen and Akin, 2010; Triantafyllidis et al., 2020). One early study from Kondraske (1994) showed that Fitts' law can be adjusted and applied to purely 2D rotational tasks, however, did not investigate combined movements.

To account for this limitation, the work of Stoelen and Akin (2010) was the first, to the best of the author's knowledge, to design an experimental setup accounting simultaneously for both translational and rotational movements. Inspired by MacKenzie (1992) shown in Equation 2.4, Stoelen and Akin (2010) proposed that by combing translation and rotation, each concept of motion must be represented by their own individual ID. Consequently, the formula of translational movements of Equation 2.4 was adjusted by replacing the otherwise target distance (A) as the numerator, with the respective rotational distance (α) and the denominator (W) indicating the target width, with the rotational tolerance (ω). In this way, an equivalent, to a great degree, ID for rotation was derived. Their model is formulated as:

$$\begin{aligned} MT_{combined} &= a + b \cdot (ID_{translation} + ID_{rotation}) \\ ID_{translation} &= log_2\left(\frac{A}{W} + 1\right) \\ ID_{rotation} &= log_2\left(\frac{\alpha}{\omega} + 1\right) \end{aligned} \tag{2.8}$$

From Equation 2.8, it is inferred that the total combined movement time it takes to point to a target ($MT_{combined}$) is dependent upon the indices of difficulty relating to translational ($ID_{translation}$) and rotational ($ID_{rotation}$) movements. From their results and as it can be further observed, in turn, from their formulation, these two IDs share the same "weight"/"scale" and are linearly correlated to MT. The implication is that both are independent concepts of motions explaining the spatial relationship of an object. Even though the simplicity of adding both terms into one ID is supported by Kulik et al. (2020),

the work of Stoelen and Akin (2010) limited their findings without the addition of either directions or inclinations, since in their experimental design pointing to the target was performed only across one line and performed in 2D space without the use of 3D displays. This limitation, however, was accounted for by Kulik et al. (2020), whereby 3D displays were used to model combined transitional and rotational movements. In the work of Kulik et al. (2020), a fairly high linear fit of $R^2 = 0.78$ when combining both translation and rotation was observed when using Equation 2.8, defined in their case as "3D Docking". Even though the findings of Stoelen and Akin (2010) were confirmed and were novel, unfortunately, Kulik et al. (2020) similarly did not account for directional movements since these were limited along one line only. In future work, Stoelen and Akin (2010) pointed out that it is important to generalise and extend a metric to three dimensions for modelling VR navigation and teleoperation. Perhaps with the hopes of quantifying the broader sensory-motor system and the entailed motor actions, these insights may aid in paving a first step towards that endeavour.

Nonetheless, it is still important to comprehensively study and infer whether rotation, in full 3D space accounting for directions and inclinations, truly shares a linear relationship with MT as translation does. The implications of these two different concepts of motion on motor difficulty (ID) are important to be researched further, to understand whether both share an equal difficulty level or whether one should consider that either concept necessitates more MT demand. Perhaps, rotation could share a polynomial or exponential relationship with MT and without properly considering all spatial variables as seen in full 3D space, conclusions are hard to draw. Consequently, it would be invaluable to further confirm their findings. With the exception of Stoelen and Akin (2010) and Kulik et al. (2020) who studied both concepts of motions, these were nevertheless limited to 2D movements across a straight line. Collectively combining all spatial variables in full 3D with 6DoFs, is scattered amongst the related work and remains a gap to this date.

The work outlined further in this thesis addresses the preceding limitations. In particular, in Chapter 4, a novel high-dimensional evaluation metric based on Fitts' law is derived, that collectively considers and combines insights of all extensions studied to this point and introduces additional insights. This metric is capable of quantifying the sensory-motor system in humans and in particular intricate high-dimensional 3D motor movements entailing 6DoFs. An overview of the characteristics of Fitts' law and all subsequent popular model extensions of it are detailed in Table 4.1a and Table 4.1b within Chapter 4.

2.2.4 Outlook, Challenges and Frontiers of Quantifying Performance

To this point, the advantages and limitations of Fitt's law have been identified. Moreover, the current extensions based on the original formulation and the important spatial aspects to consider to truly be able to generalise the full 3-dimensional domain have been outlined. However, there are additional challenges to address prior to unifying all spatial associated variables as seen in higher dimensions and covering the full 3D domain entailing higher DoF movements.

Consequently, in this section, the challenges and frontiers that need to be addressed are outlined, to consider generalising the original formulation of Fitts to higher dimensions and more intricate motions. This section will furthermore go beyond those challenges and frontiers and report the outlooks of what it entails to standardise, in a broader context, human motor performance. Finally, in Table 2.1 a summary of potentially important research directions, aims and questions is formed. The author of this thesis hopes to present potentially important directions and challenges that researchers are expected to encounter for deriving a standardised evaluation tool that can quantify human motor performance, and that can furthermore be generalised to a plethora of different settings.

Implications of Depth Perception

Foremost, as identified in Section 2.1, vision is the dominant sense in humans, accounting for approximately 70% towards the sensory contribution of all senses according to Heilig (1992). Consequently, it should come as no surprise that depth-associated distances are one of the key aspects, as well as limitations, one should consider when concerned with motor movements and even more so when attempting to quantify these (Sigrist et al., 2013; Lampton et al., 1995). The implications of depth perception are of further importance when employing MR-based technologies for the assessment of motor movement (Barrera Machuca and Stuerzlinger, 2019).

The main limitation of applying Fitts' model in 3D pointing, especially with the use of MR technologies, or any human predictive performance model for that matter, is mostly attributed to the lack of properly accounting for depth associated variables (Hong and Kang, 2015; Lubos et al., 2014; Batmaz et al., 2019). In particular, findings from numerous studies attribute this limitation due to distance overestimation in VEs (Lampton et al.,

1995; Renner et al., 2013; Witmer and Kline, 1998; Swan et al., 2015). The work of Renner et al. (2013) even estimated that the discrepancy of depth estimation differed with the real target to almost 74% of their true distance. Other studies have suggested a few mitigation approaches to overcome depth limitations, one of which includes increasing the display resolution of 3D displays to provide higher representations of depth and limit to some extent distance overestimation (Godse et al., 2019; Kenyon and Ellis, 2014). Further studies concerned with measuring motor performance via Fitts' model confirmed that movements along the depth axis (away or towards the user) appear to be significantly more difficult than left-to-right movements, which appears to be supported by earlier works of Murata and Iwase (2001); Cha and Myung (2013); Schofield (1976); Barrera Machuca and Stuerzlinger (2019). Nevertheless, to what extent these depth implications affect motor difficulty (by extent MT) necessitates further research. For instance, whether depth variables share a linear or non-linear relationship with motor difficulty, especially when deriving formulations based on Fitts' law, requires further investigation.

Human Factors and Associated Complexities with Experimental Designs

By definition, Fitts' prediction model, attempts to quantify the performance of human movement timings based on spatial settings. A plethora of different human factors influence human performance models, and it's essential to consider them. Some of these factors include yet are not limited to; personality-related factors such as age, visual health and previous exposure to certain technologies, amongst others (Hess et al., 2015; Zaroff et al., 2003; Sacau et al., 2008; Menchaca-Brandan et al., 2007). Moreover, elements such as tiredness, concentration, being under the influence of stimulants and more, may have a determining effect on the derivation of a performance model, especially when attempting to standardise such metrics (McGlynn and Rogers, 2017). More broadly, these associated complexities can also influence any conclusions drawn when attempting to establish a correlation between sensory feedback and generating motor actions, as explored in Section 2.1.

From a practical standpoint, accounting for all these factors in a formulation can be notably challenging. Nonetheless, an appropriate mitigation strategy that could be utilised would be during initial participant recruitment. For instance, during recruitment, a set of selection criteria could be set forth to limit additional human-factor-related variables that could be introduced and to retain consistency (Drewes, 2010). In particular, participants in a study could be evaluated on whether they are tired or are under the influence

of certain stimulants, such as coffee, with those not meeting the selection criteria to be subsequently excluded from the analysis. Nevertheless, related studies so far do not include, to some extent, a clear definition of the state of the participants which would aid subsequent inter-study comparability amongst the related work (Murata and Iwase, 2001; Cha and Myung, 2013; Kulik et al., 2020).

Other important factors to take into account are the specific experimental designs that authors consider as part of their studies when deriving performance metrics. For instance, deriving a human performance model under a single experimental setting is likely going to be insufficient as the authors are likely at the risk of observing and tuning the model parameters to fit that particular study's data. Ultimately, if a human performance model is only applicable to very specific applications or experiments, its relevance can become questionable and is likely at risk of losing generalisability. To address to a great extent some of these limitations effectively, one could consolidate all extensions of a specific predictive model, such as Fitts' law, and then evaluate and contrast them under the umbrella of a singular 3D experimental design context. This would notably help support the generalisability or standardisation arguments made by scholars (Triantafyllidis and Li, 2021a).

Complexities Associated with Evaluating Performance Models

An additional and important part of the challenges and frontiers in deriving a standardised human motor performance metric are the current evaluation methods used, which limit inter-study comparability and conclusions (Drewes, 2010; Gori et al., 2018; Triantafyllidis and Li, 2021a).

One of the most popular approaches in evaluating Fitts' model and its extensions, is the utilisation of the coefficient of determination (r^2 or R^2). This coefficient is used to assess the correlation between the ID and MT, which when visually depicted would be drawn on the x and y-axis respectively. An r^2 value closer to 1 indicates a "stronger" linear relationship between these two variables, suggesting a "better" fit. Conversely, an r^2 value nearing 0 implies a "weaker" correlation and by extent, a "lesser" fit. Via this evaluation method, an extension is deemed to be "superior" when evaluated against other models when there is a better correlation. Nevertheless, there are numerous disadvantages to merely using the (r^2), which is supported by current literature (Drewes, 2010; Gori et al., 2018). The shortcomings of just relying on the r^2 are outlined in this section.

The major limitation of merely relying on the coefficient of determination is that it is highly dependent upon the number of data points recorded as evidenced by Triantafyllidis and Li (2021a). In particular, the more data pairs there are in the evaluation, the correlation will usually be lower. This disadvantage can be exploited by deriving a small number of IDs and a significant amount of repetitions of the same tasks to achieve relatively "easily" a good resolution. The work of Drewes (2010) exploited this limitation of the r^2 in a single vs multiple times repeated click-the-target experiment.

Inspired by the findings of Drewes (2010), the author of this thesis aimed to reproduce their experiment to verify these results and applied it in a preliminary setting consisting of pointing to a target task, entailing a VR-based experiment with more levels of repetition. In this experiment, the author conducted a simple pointing task with four target sizes *(**W** = 5, 7.5, 10, 12.5 [cm])* and four target separations *(**A** = 12, 24, 36, 48 [cm])*, accounting to a grand total of 16 distinctive tasks. When applying Fitts' law as detailed in Equation 2.1, a total of 16 ID were calculated, ranging from 0.941 to 4.26 bits. A total of 20 repetitions were made by a single participant.

In Figure 2.1 it can be visually observed how 1, 5, 10 and 20 repetitions affect the r^2 fitting values of the simple 3D pointing-to-the-target experimental task the author conducted. From Figure 2.1 it can quite clearly be inferred that as the repetitions increase, significantly higher r^2 values are in turn rendered.

Unfortunately, to date, most related literature use this evaluation approach for the most part (Batmaz et al., 2019; Murata and Iwase, 2001; Cha and Myung, 2013; Stoelen and Akin, 2010; Kulik et al., 2020). While it does indeed provide an overview of how well a model explains the data, researchers would be advised to report the values of the constants as part of Fitts' original model in addition to the respective standard error as shown in Figure 2.1. For instance, solely stating that a model has a correlation of $r^2 = 0.906$ should be avoided, and instead include a statement similar to: *"Fitts' original formulation showed a correlation of* $r^2 = 0.906$ *with MT = 0.66 (0.07) [0.52, 0.79] + 0.27 (0.02) [0.21, 0.31]* $\cdot$ *ID"*. In this type of statement, the terms would represent the a and b constants respectively, reporting the standard error in addition to the lower and upper bound Confidence Intervals (CI) (typically set at CI=95%) that can be reported through regression analysis. Moreover, it is advisable that when employing the coefficient of determination (r^2), to avoid direct inter-study comparisons and rather include existing metrics as part of the same experimental evaluation with an identical number of repetitions.

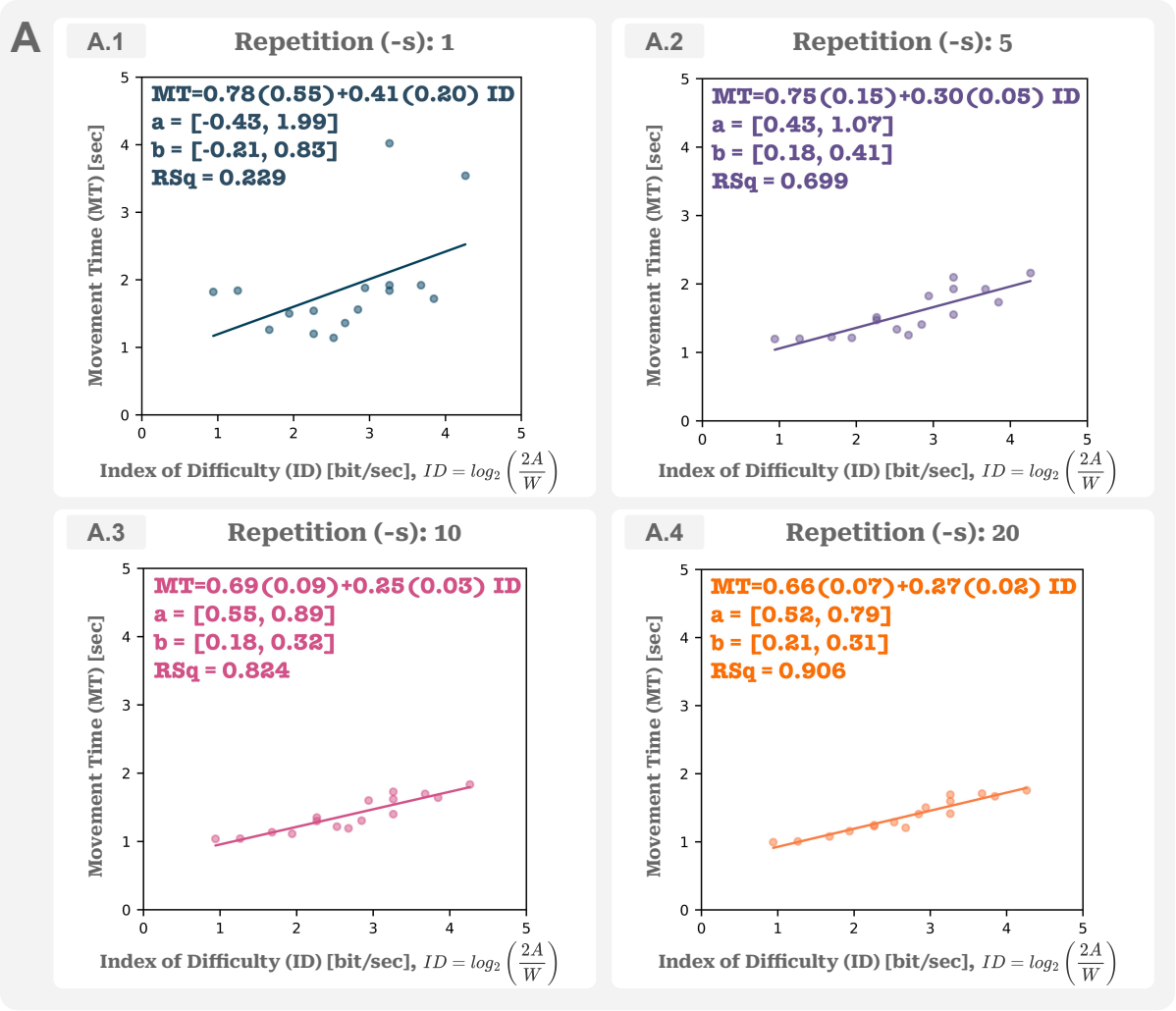

Figure 2.1: **Representative example of the effects of repetitions and data points on regression.** The figure depicts an illustrative example of how different repetitions ($A.1 - A.4 \in \{1, 5, 10, 20\}$) affect linear regression and in particular the coefficient of determination – R^2. These data points were acquired from a simple point to the target task utilising Fitts' original law from Equation 2.1 (Fitts, 1954). This task serves as an example of how an increased number of repetitions leads to higher R^2 results based almost exclusively on the number of data points. Notice the notable increase of the (R^2) for repetition 1 (A.1), 5 (A.2), 10 (A.3) and 20 (A.4) corresponding to $R^2 = 0.229$, $R^2 = 0.699$, $R^2 = 0.824$ and $R^2 = 0.906$ respectively. It can be inferred that this significant effect of the data points on the R^2, limits the ability to assess and determine which model, in the context of measuring or capturing human motor performance, is "superior" over others. An appropriate mitigation strategy that can minimise to some extent these shortcomings, is to introduce an identical number of repetitions for all compared models. Moreover, this figure serves as an example of a more suitable and helpful overview of how the model fitting results between the ID and MT should be. In particular, one should report the full line equation $MT = a(SE) + b(SE) \cdot ID$, with ($SE$) representing the standard error and reporting the lower and upper bound CI of the constants (CI=95%).

Multitude of Input Technologies and Effects of Physical Quantities

Previous work and more specifically related to Fitts' model, have indicated that the multitude of input technologies that can be utilised, notably influence the formulation (MacKenzie et al., 1991; Wobbrock et al., 2008; Forlines and Balakrishnan, 2008; Wobbrock et al., 2011b). Proposing a full 3D extension of a performance model may furthermore be aggravated by this factor alone. It should come as no surprise that numerous models may need to be derived, each for a very specific input technology (Triantafyllidis and Li, 2021a). In the context of using Fitts' law, perhaps the introduction of an additional constant specific to the input device may be of necessity (Triantafyllidis and Li, 2021a; Barrera Machuca and Stuerzlinger, 2019).

In the context of robotic manipulation and the physical interaction of objects, the entailed physical properties and grasping types one can use may have a profound effect on a formulation. While Fitts' law was originally designed to be used exclusively for pointing tasks, authors have extended it to manipulation, i.e. with the addition of physically grasping items as observed in the work of Triantafyllidis et al. (2021). Nevertheless, other physical phenomena such as friction or gravity, may have determining effects, necessitating additional investigations (Teather and Stuerzlinger, 2007). Moreover, complex hand end-effectors that closely resemble anthropomorphic features can be challenging to incorporate in a formulation, and metrics that measure the quality of grasping may offer promising alternatives (Zheng, 2013; Ferrari and Canny, 1992).

Lastly and closely related to the multitude of anthropomorphic and non-anthropomorphic hand types that can be used, the grasping types humans can initiate with an object may prove to be determining factors in the manipulation domain. For instance, the Southampton Hand Assessment Procedure (SHAP) is a clinically validated hand function test, consult Light et al. (2002) that reports six in total distinctive grip classifications: precision-tip, lateral, tripod, spherical, power as well as extension grasping. The work of Kumar and Todorov (2015), studied the effects of the different SHAP grasping types in a VR system for hand manipulation, yet not with the aim of deriving a metric. Nevertheless, one could consider their study as the means of building an apparatus that can accommodate the different SHAP grasping types and investigate their effects on motor movement for deriving an appropriate evaluation tool (Triantafyllidis and Li, 2021a).

Table 2.1 summarises potentially important research questions and directions with the future aim of deriving a standardised evaluation tool.

Table 2.1: **Potentially important research questions, directions and paths for the derivation of a robust performance metric.** These considerations may aid in solidifying a standardised metric, generalised for a plethora of different settings.

Research Questions and Considerations to Explore	**Sources**
How do different types of input technologies and devices affect a given model formulation? For Fitts' law, can it be accounted for by implementing an additional term or does it suffice to control existing terms with additional constants?	Burno et al.; Sasangohar et al.; McGee et al.
Spatial arrangements, directions & inclinations, appear to affect motor movement difficulty according to the related work. However, can existing findings be confirmed and extended in combined translational and rotational movements? Potentially of notable importance due to limited focus.	Murata and Iwase; Cha and Myung; Barrera Machuca and Stuerzlinger; Vaughan et al.
Is the simplicity of adding two indices of difficulty of translation and rotation plausible? Can rotational movements be modelled in the same way as translational movements and simply adding these terms derive a combined movement ID? For translational tasks, from the related work, it is supported that for Fitts' law the derived IDs share a linear relationship with MT, is that the case for the less explored rotational counterpart? Do rotational movements perhaps share a polynomial or even exponential relationship with MT? Existing work has, to some extent, explored how combining these two separate types of motions can be achieved. However, to date, findings are limited to movements across one line only with no directions or inclinations and mostly in 2D space.	Stoelen and Akin; Kunert et al.; Wang et al.
With the above in mind, researchers are advised to explore other extensions when it comes to combined movements. Both Stoelen and Akin, as well as Kulik et al. work, extended the formulation of MacKenzie exclusively. Hoffmann's or Welford's should also be explored with their differences reported.	Stoelen and Akin; Kulik et al.; Kunert et al.; Wang et al.; MacKenzie; MacKenzie and Isokoski; Hoffmann et al.; Welford
Depth estimation in VEs differs from that of the real world, with users overestimating the distances. Can the work of Barrera Machuca and Stuerzlinger, which is to the best of the author's knowledge the only work that accounted for the depth variable and added to Fitts' (MacKenzie's) extensions, suffice? With the spike in MR technologies, verification of the above should be prioritised and accounted for.	Barrera Machuca and Stuerzlinger; Lin and Woldegiorgis; Durgin et al.; Witmer and Kline; Lampton et al.; Renner et al.
Models should not solely be evaluated on the coefficient of determination (r^2). As explored from the preliminary findings presented in Figure 2.1, it can be inferred that purely relying on the r^2 fitting can introduce limitations as part of the evaluation procedure. Mitigation strategies include: (i) to report the full line equation of MT, entailing constants, standard error and CI (such as at CI=95%), as well as (ii) all relevant formulations under the same study to retain consistency in terms of repetitions and experimental design.	Drewes; Gori et al.
Can most, if not all, human factors be accounted for? Tiredness and concentration for example appear to be key elements affecting notably human motor performance according to the literature. However, modelling these factors can be challenging. A possible mitigation would be for researchers to account for these factors during participant recruitment and retain consistency.	de Grosbois et al.; Drewes; Rozand et al.; Zeng et al.; Poletti et al.; Thumser et al.
Most related work explored Fitts' law purely with visual stimuli. How do other modalities affect object interaction, in particular haptic or auditory feedback? Can one reduce potentially impaired amounts of depth perception by exploring different sensory combinations?	Teather and Stuerzlinger; Liu et al.; Gemperle et al.; Brickler et al.; Triantafyllidis et al.; Bunt et al.; Amedi et al.; Turk; Quek et al.; Vertegaal
Manipulation entails different physical quantities such as gravity, contact points and more. These could prove to be invaluable, especially in simulations entailing high-fidelity physics (Sim2Real). Can these properties be accounted for? Furthermore, how do different hand types with different contact points, degrees of freedom and even grasping classifications affect a particular model?	Triantafyllidis et al.; Zheng; Burstyn et al.; Zhai et al.; Wang et al.; Kunert et al.; Thumser et al.
What are the differences that arise in pointing versus manipulation types of interactions? Can these two different types of interactions be modelled under one unified formula?	All of the above sources.

2.3 Related Work III: Biologically-Inspired Embodied Intelligence and Hierarchical Learning

Until this point, it can be inferred that by virtue of the highly intricate sensory-motor system, humans are able to interact with their surrounding environment with what appears to require minimal effort. Section 2.1 and Section 2.2 explored how the sensory-motor system can be harnessed and in particular how sensory states can contribute towards generating meaningful motor actions, which can thereafter potentially be measured and quantified.

Understanding the underlying biological mechanisms of the sensory-motor system in humans could illuminate ways of bridging these biological capabilities closer to embodied intelligence. Inspired by this notion, this section reviews existing machine learning approaches, with an emphasis on aligning human skills to embodied intelligence, from a neuro-biological perspective. In particular, this section will focus on intricate challenges intelligent agents are faced with, most notably sparse rewards and long-horizon tasks, in the narrower context of robotic manipulation, with effective ways of overcoming said challenges. With the hope of deriving embodied intelligence that can exhibit intelligent, robust and adaptive behaviour as to address and meet intricate environmental challenges, drawing inspiration from biological insights may serve as the foundation for realising this.

2.3.1 Biological Abilities and the Role of AI in Embodied Intelligence

Humans are capable of interacting with their surrounding environment and meeting highly intricate environmental demands seamlessly (Triantafyllidis and Li, 2021a; Ashe et al., 2006; Ortenzi et al., 2019). These astonishing capabilities are in large part due to the highly intricate sensory, cognitive and dexterous skills that humans inherently possess (Triantafyllidis et al., 2020; Billard and Kragic, 2019).

In stark contrast to the above, robots, viewed as agents of embodied intelligence, are currently and to a large extent, unable to mimic too closely these physical interactions (Tee et al., 2022; Billard and Kragic, 2019). Solving notably complex sequential tasks over a long horizon remains an ongoing challenge in this domain (Davchev et al., 2022; Fox et al., 2019). Nevertheless, embodied intelligence and in particular robots can perform repetitive tasks with high precision, provided these are confined to predefined and specific

tasks (Triantafyllidis et al., 2020; Zhang et al., 2021). Some of these tasks range from picking and placing (Zhang et al., 2018b; Triantafyllidis et al., 2020), swing-peg-in-hole (Chebotar et al., 2019; Lee et al., 2019), catching in-flight objects (Schill et al., 2015), insertion (Lee et al., 2019; Schoettler et al., 2020) to solving tasks as complex as a Rubik's cube (Andrychowicz et al., 2020). Nevertheless, when it comes to solving a sequence of multiple tasks that vary in complexity, notable challenges arise (Zhang et al., 2021).

The use of analytical models to program a robotic system for predetermined tasks may result in sub-optimal outcomes due to these models often oversimplifying real-world dynamics. This, in turn, demands costly computation in real-time and is unable to accommodate for changes in physical properties as well as dynamically changing environments as seen in the work of Triantafyllidis et al. (2023b). However, recent developments in AI and Machine Learning (ML) offer a promising solution for enhancing robot learning and embodied intelligence (Lee et al., 2019; Zhang et al., 2018b,a; Yang et al., 2020; LeCun et al., 2015).

2.3.2 Reinforcement Learning Algorithms – Trial and Error

Deep learning algorithms and in particular RL approaches offer compelling solutions in overcoming the inherent and prevailing challenges of pre-programming robots with embodied intelligence (LeCun et al., 2015). Unlike traditional programming especially when employed in embodied intelligence, RL offers a promising alternative due to its innate capacity to learn from interactions with its environment and progressively adapt to it (Lee et al., 2019; Zhang et al., 2018b). Within the RL setting, agents strive to learn a set of policies by maximising their environmental returns, known as extrinsic rewards. The RL paradigm also closely mimics a biological perspective, whereby even human infants in the early stages of their lives must learn unlabelled data to infer useful representations and establish meaningful patterns (Zaadnoordijk et al., 2022; Saxe et al., 2021).

Such adaptive learning offers a more dynamic and flexible way of meeting intricate environmental demands, commonly encountered in embodied intelligence and especially in the intricate dimension of robotic tasks such as manipulation (Triantafyllidis et al., 2023b). Consequently, RL has seen a myriad of applications, especially in robot learning due to its capacity to evolve and adapt to environmental interactions and demands (Lee et al., 2019; Zhang et al., 2018b,a; Yang et al., 2020). Consequently, as one can infer, instead of prescribing and pre-programming a system's every move, utilising RL instead enables

the machine to grasp, adapt, and refine its actions, closely mimicking how humans learn from experiences as seen from a biological standpoint, as suggested by Zaadnoordijk et al. (2022).

Amongst the relevant studies, the Proximal Policy Optimization (PPO) from the works of Schulman et al. (2017) and Soft Actor-Critic (SAC) from Haarnoja et al. (2018), are the most commonly used RL algorithms. PPO is an on-policy algorithm while SAC is off-policy. While PPO is on-policy and generally less sample efficient than the off-policy algorithm SAC, PPO is still less prone to instabilities and typically necessitates less hyperparameter tuning than SAC (Haarnoja et al., 2018; Schulman et al., 2017; Gu et al., 2017). For these reasons, PPO is chosen as one of the main RL algorithms in subsequent chapters of this thesis.

2.3.3 Supervised and Imitation Learning – Priory Knowledge

While RL algorithms are promising, these are still limited and do not perform, to a great extent, too well with intricate tasks that are challenged by sparse rewards and long horizons, whereby environmental feedback may not be readily apparent, which ultimately exacerbates the relevance of the exploration-exploration trade-off (Koganti et al., 2018; Ding et al., 2019; Zaadnoordijk et al., 2022). The major drawback of RL algorithms is their primary lack of prior knowledge about a given task. Typically, these methods learn to solve a new task by generating their own experience, without any initial guidance, necessitating learning from scratch, which can be highly inefficient if the task is notably intricate (Schaal, 1997; Zador, 2019). Moreover, this learning process of starting from scratch can involve millions of state transitions and can often require several days to complete (Thor and Manoonpong, 2022; Yang et al., 2020).

Imitation Learning (IL) presents itself as an alternative approach to RL, based on the premise that humans possess prior knowledge when learning motor tasks, rather than starting from scratch (Goldberg, 2019). IL, also known as Learning from Demonstration (LfD), involves the provision of expert demonstrations to enable agents to emulate and imitate a demonstrated behaviour. This approach has shown promise in intricate and dexterous robotic tasks that would be difficult or even impossible to attain using conventional RL approaches, which require extensive exploration and careful reward specification (Zhang et al., 2018b; Schaal, 1997; Koganti et al., 2018). Hand-engineering such specific rewards that provide continuous feedback in RL may overcome some of the exploration

challenges faced, yet still necessitate notable engineering efforts and are rendered highly task and case-dependent, notably limiting generalisation to other tasks and environments (Pathak et al., 2017; Du et al., 2023; Amodei et al., 2016; Triantafyllidis et al., 2020).

Typically, IL and LfD approaches rely on expert demonstrations from humans. However, some alternatives to expert human demonstrations can be utilised, such as conventional trajectory optimisation (Zhang et al., 2018b; Levine and Abbeel, 2014) or RL approaches (Mnih et al., 2015; Schulman et al., 2015; Mnih et al., 2016). Nevertheless, these methods generally require carefully designed costs or rewards and notable interaction time between the robot and the environment, while at the same time not leveraging the invaluable cognitive insights and scene understandings that humans possess (Billard and Kragic, 2019; Triantafyllidis et al., 2020; Triantafyllidis et al., 2023b).

Behavioural Cloning

Behavioural Cloning (BC) is one of the primary IL algorithms used in related work, which involves performing supervised learning on the policy using a set of demonstrated state-action transitions. Once trained, the learned policy can be used to generate actions for new states, which ideally should be similar to the actions taken by the experts in those states. This approach has shown fairly good success in robotic tasks as supported by Zhang et al. (2018b); Pastor et al. (2009); Ratliff et al. (2007); Fox et al. (2019), constituting it a promising approach for warm-starting a policy without having to rely on conventional RL approaches that start from scratch. However, BC has several limitations when used in isolation, some of which include the lack of exploration, the limited robustness towards new non-encountered states as well as the dependence on large, near-optimal demonstration datasets according to Ho and Ermon (2016).

Foremost, naively mimicking the expert demonstrations via BC is prone to lead to problematic performance when the agent visits states not encountered in the demonstrations due to compounding error in covariate shifting, which in turn results in the need for large amounts of demonstration data (Ho and Ermon, 2016; Ross et al., 2011). Moreover, the need for a large demonstration dataset by a human can eventually lead to operator fatigue resulting in degraded performance over time (Triantafyllidis et al., 2020; Triantafyllidis et al., 2021).

Looking at the matter from a biological perspective, relying entirely on an expert to acquire new skills without any additional learning strategies seems to be misguided on

a fundamental learning level (Zaadnoordijk et al., 2022; Zador, 2019; Saxe et al., 2021). Although it may seem counter-intuitive to start learning a new skill from scratch, Zaadnoordijk et al. (2022) offered an appropriate analogy from a biological perspective, whereby trial and error is a vital part of the early lives of humans: *"Human infants are in many ways a close counterpart to a computational system learning in an unsupervised manner, as infants too must learn useful representations from unlabeled data"*. Switching back to a learning perspective, this could perhaps imply that learning in its core, should not entirely and solely be dependent upon imitating an "expert". Instead, further exploration beyond "naively copying others" should be allowed and in this way still draw inspiration from a neuro-biological perspective (Zador, 2019; Saxe et al., 2021). This could imply that a balance between exploration (trial and error) and exploitation (imitation), could result in a positive learning procedure while also drawing inspiration from a neuro-biological standpoint.

Generative Adversarial Imitation Learning

To overcome the limitations of BC, an alternative option is to use Inverse Reinforcement Learning (IRL), which involves inferring the reward function from underlying observed demonstrations to explain them in terms of near-optimal behaviour (Ho and Ermon, 2016; Abbeel and Ng, 2004; Finn et al., 2016). Generative Adversarial Imitation Learning (GAIL) is a popular IRL algorithm that utilises Generative Adversarial Network (GAN)s (Ho and Ermon, 2016). In GAIL, a second Neural Network (NN) referred to as the discriminator is used to differentiate between sets of trajectories generated by the agent or the expert and the less their divergence is, the greater the reward in the IL framework (Ho and Ermon, 2016).

2.3.4 Hierarchical Learning Architectures

While fusing deep RL and IL approaches has the potential to endow robots with some of the capabilities exhibited by human cognition, there are still notable challenges to be solved. When approaching large-scale problems, the most notable frontier is sequential of long-horizon dexterous manipulation tasks (Marzari et al., 2021; Fox et al., 2019).

The Limitations of Monolithic Single Neural Networks – Training a single monolithic neural network to learn and solve notably intricate tasks can be significantly difficult and challenging due to: (i) long horizon problems, where the computational

complexity of approximating a policy is very high, (ii) the variability of the task typically entails many sub-tasks and sequences, as well as (iii) the sample complexities associated with complex dexterous robotic tasks (Le et al., 2018; Behbahani et al., 2019; Fox et al., 2019; Davchev et al., 2022). Further aggravating the use of single NNs and utilising these for long-horizon problems is the necessity of completing a high-level end goal with environmental feedback potentially not readily available. For long sequential tasks, this end goal can be notably intricate, often contingent upon the successful completion of all involved sub-tasks and usually, in a particular sequential order (Marzari et al., 2021). In addition, even if such sub-tasks are decomposed into smaller, more approachable problems (Rajeswaran et al., 2018; Marzari et al., 2021), these can still vary notably between each other, limiting learning potential due to limited task inter-relation (Liu et al., 2018).

Hierarchies can somewhat alleviate some of the above complexities. Using Hierarchical Learning (HL), either coupled with RL or IL, is a promising strategy for addressing the issues mentioned above and can help reduce some of the difficulties (Frans et al., 2018; Merel et al., 2019a,b; Yang et al., 2020). HL has several advantages for multitasking or other intricate tasks associated with sparse rewards (Davchev et al., 2022), as it enables the breakdown of tasks into smaller sub-tasks, often referred to as "skills" (Fox et al., 2019). To tackle complex robotic tasks, a MoE architecture represents itself as a popular approach. In this architecture, multiple task-specific experts, often referred to as "skills", are trained for a given sub-task while supervised and managed by a primary gating network utilising these as macro-actions. These types of architectures have been successfully applied in computer graphics and animation (Zhang et al., 2018a; Peng et al., 2019) as well as robotics (Mülling et al., 2013; Fox et al., 2019; Yang et al., 2020).

Hierarchical Reinforcement Learning – When HL is coupled with RL, commonly referred to as a Hierarchical Reinforcement Learning (HRL) framework, notably intricate robotic tasks can be solved, especially when coupled with a MoE architecture (Yang et al., 2020). However, even in such complex architectures, there are still some notable challenges. HRL relies fundamentally on the RL paradigm and is adversely affected and unable to cope with sparse rewards, complex planning tasks or the integration and experience of prior knowledge to solve tasks (Fox et al., 2019; Marzari et al., 2021).

Hierarchical Imitation Learning – To overcome the limitations of RL and HRL, IL can be used as an alternative approach that leverages expert demonstration data to

facilitate a faster training process by utilising said data as a warm-start for the policy. When hierarchical learning is combined with IL to create a Hierarchical Imitation Learning (HIL) framework, it is rendered easier to differentiate and isolate between specialised experts and for a human to acquire specialised skills in a teacher-student relationship. This is in virtue of the breakdown of sub-tasks in a hierarchical control policy, rendering the acquisition of these skills notably more straightforward (Le et al., 2018; Fox et al., 2019, 2018). This is in contrast to RL and HRL, which rely on trial-and-error learning and can be adversely affected by sparse rewards and complex planning tasks (Fox et al., 2018, 2019). This, in turn, allows the utilisation of HIL-based architectures for notably more intricate, longer horizon sequential tasks as opposed to HRL-based ones (Fox et al., 2018, 2019).

Nevertheless, when it comes to robotic manipulation, the utilisation of MoE architectures that are trained using HRL or HIL techniques is presently limited in the latest advancements of state-of-the-art (Rajeswaran et al., 2018; Marzari et al., 2021). Based on prior work that introduced ensemble techniques in robot locomotion (Yang et al., 2020) and human-centred teleoperation (Triantafyllidis et al., 2021; Triantafyllidis et al., 2020), the motivation for the third contribution of this thesis is built, to explore a new method of IL using human-demonstrated tasks developing a suitable MoE architecture in the domain of robotic manipulation. In this way, a new approach can be derived with the potential of extending beyond demonstrated behaviour by fusing multiple learning paradigms and enabling, in turn, more complex and adaptive manipulation tasks.

More closely related to manipulation, the work of Marzari et al. (2021) employed an HRL strategy that involved three experts in training a robotic gripper to perform a given task by (i) *approaching*, (ii) *manipulating* and (iii) *retracting* from it. Although their outcomes were authenticated against IL (BC) and indicated greater success rates ($>$ 90%) in comparison to utilising RL, the tasks they investigated remained fairly simplistic, assuming non-sequential tasks with rather short horizons. Consequently, their assessment was limited to three experts solving fairly simplistic pick-and-place tasks. As it will be presented in Chapter 5, a novel method that was derived from this PhD work not only shows that a *single expert* is capable of solving picking and placing tasks but when combined with other experts specialising in rather high-level sub-tasks compared to the work of Marzari et al. (2021), notably intricate of long-horizon sequential tasks commonly seen in robotics can be addressed.

In particular, in (Chapter 5), a novel biologically-inspired hierarchical learning architecture is derived that fuses multiple learning paradigms, tailored for embodied intelligence challenged by notably intricate, sequential and long-horizon robotic manipulation tasks.

2.4 Related Work IV: Language-Guided Learning

To this point, it is identified that humans possess unmatched motor dexterity and are capable of meeting highly intricate environmental demands by virtue of their cognitive abilities (Triantafyllidis and Li, 2021a; Ashe et al., 2006; Ortenzi et al., 2019). Consequently, Section 2.1 and Section 2.2 of this chapter were dedicated to understanding the human sensory-motor system and in particular how it can be harnessed and quantified. Thereafter, switching from a biological to an embodied intelligence perspective, Section 2.3 sought to identify how human skills can be leveraged in intricate settings, as to align human capabilities closer to AI, inspired by a biological standpoint as with the works of (Triantafyllidis et al., 2023b).

Given the aforementioned, this final section (Section 2.4) seeks to identify how recent advances in LLMs and in particular how eliciting their context and common sense aware reasoning (Du et al., 2023; Yu et al., 2023; Wei et al., 2023), can aid existing learning settings challenged by sparse rewards and long-horizons. Ultimately, the author of this thesis is seeking to create a capable biologically-inspired approach that leverages human cognition that could further be enhanced and solidified by language-guided reasoning. This motivation is with the aim of offloading certain amounts of human supervision, especially for higher-level scene understandings. Consequently, this section attempts to identify how language-guided insights may aid and enhance biologically inspired embodied intelligence challenged by intricate settings, entailing a plethora of environmental and task-specific demands.

2.4.1 Reinforcement Learning – Trials, Errors and the Pursuit of Elusive Sparse Extrinsic Rewards

As identified in the related work, pre-programming robots as embodied intelligence via analytical models can be sub-optimal due to the overall oversimplification of modelling real-world dynamics (Triantafyllidis et al., 2023b). A promising alternative in overcoming this challenge is deep learning algorithms (LeCun et al., 2015) and in particular deep RL

algorithms. RL presents itself as a promising basis with its capacity to learn from interactions with the environment (Lee et al., 2019; Zhang et al., 2018b), drawing inspiration from a biological perspective whereby even human infants in their early lives must learn representations from unlabelled data according to Zaadnoordijk et al. (2022); Saxe et al. (2021).

In RL in particular, agents aim to learn a set of policies by maximising their returns from the environment. These returns are known as extrinsic rewards. Such approaches have shown promising applications in robot learning due to their capacity to adapt and learn from the agent's interaction with its environment (Lee et al., 2019; Zhang et al., 2018b,a; Yang et al., 2020). When such extrinsic rewards are provided in an immediate or continuous manner, RL policies tend to perform very well (Triantafyllidis et al., 2023b). However, in many RL settings and most notably in real-world robotic cases, these extrinsic rewards can be all but immediate. In particular, extrinsic rewards tend to be mostly (i) **sparse**, providing non-immediate feedback from the environment, which exacerbates the exploration-exploitation trade-off (Zaadnoordijk et al., 2022; Koganti et al., 2018; Ding et al., 2019), (ii) **miss-aligned** (Schäfer et al., 2022) or (iii) necessitate very careful **crafting and tuning** which limits generalisation as these ultimately become task-specific and use-case dependent (Schäfer et al., 2022; Du et al., 2023). Even hand-crafting specific dense rewards that provide continuous feedback may initially overcome some of the exploration challenges, these still necessitate notable engineering efforts and are rendered highly case dependent notably limiting generalisation to other settings (Pathak et al., 2017; Du et al., 2023; Amodei et al., 2016).

Ultimately, for agents to randomly reach and attain a sparse goal is highly unlikely due to the increased need for exploration (Schäfer et al., 2022). This, in turn, necessitates prohibitively long resource-intensive learning sessions without still a guarantee of convergence to the sparse goal (Thor and Manoonpong, 2022; Yang et al., 2020). This is especially the case in long-horizon sequential robotic manipulation tasks whereby environment feedback may not be readily apparent (Davchev et al., 2022; Triantafyllidis et al., 2023b).

2.4.2 Intrinsic Guidance – The Compass Navigating Through the Maze of Reinforcement Learning

To tackle environments with sparse rewards, a prominent category of exploration techniques used are intrinsically-motivated rewards which foster and encourage exploration

(Schäfer et al., 2022; Pathak et al., 2017; Burda et al., 2019; Taiga et al., 2020; Bellemare et al., 2016). Intrinsic rewards (r^i) are computed and added to the existing extrinsic reward from the environment (r^e). Given these two rewards, the total combined reward (r^c) is represented as $r^c = r^e + \lambda^i \cdot w^i \cdot r^i$, with λ^i representing the intrinsic scaling factor. Moreover, over time, r^i is decayed, usually via a linearly decaying weight (w^i), by which point the agent should converge towards the extrinsic reward, transitioning from exploration to exploitation (Schäfer et al., 2022). There are two common categories of intrinsic rewards amongst the related work, (i) **count-based** and (ii) **prediction-based** intrinsic reward methods.

Count-based exploration methods are traditional intrinsic methods that incentivise agents to visit rarely encountered states. Essentially, the agent receives higher intrinsic rewards for those states that have been visited less frequently. While count-based exploration techniques are useful in small, discrete state spaces, these can be less effective in high-dimensional state spaces (Bellemare et al., 2016). This can especially be challenging during intricate robotic manipulation tasks entailing long horizon and sequential operations as seen in the works of Triantafyllidis et al. (2023b).

Prediction-based methods, as the name suggests, is a separate approach to defining intrinsic rewards utilising predictions in the environment. Recent popular prediction-based methods that have been proposed are the Intrinsic Curiosity Module (ICM) (Pathak et al., 2017) and the Random Network Distillation (RND)(Burda et al., 2019). With ICM, a prediction error is utilised to promote the agent to explore new and unfamiliar states. While encouraging exploration, ICM can nonetheless overemphasise noisy state transitions due to prediction error, potentially leading the agent to explore non-relevant to a given task state transitions (Pathak et al., 2017; Taiga et al., 2020). Contrary to ICM, RND generates intrinsic rewards by comparing feature representations of the agent with those of a fixed random network. Unfortunately, RND's static random network might not always capture evolving complexities in dynamically changing environments such as those encountered in physics-based interactions and most prominently encountered with robotic manipulation (Triantafyllidis et al., 2023b; Burda et al., 2019; Taiga et al., 2020). Further research and utilisation of these methods on intricate robotic manipulation tasks is necessitated to underline their potential as well as shortcomings.

It is inferred that while intrinsic rewards can promote exploratory potential and appear to be promising means of guiding agents in environments that are challenged by sparse

rewards, these methods may nonetheless explore noisy states and may focus on states beyond the relevance of the task at hand (Schäfer et al., 2022; Pathak et al., 2017; Burda et al., 2019; Taiga et al., 2020; Bellemare et al., 2016). Perhaps, by fusing the exploratory potential of intrinsic rewards, in the form of contextual insights from LLMs, a more directed and efficient exploration strategy can be devised.

2.4.3 Language-Guided Reinforcement Learning – Eliciting Reasoning for Exploratory Potential

An alternative to providing useful context as well as common-sense aware reasoning that could aid agents in exploring states that are directly relevant to the task at hand is LLMs. LLMs are capable of providing textual insights entailing context-aware reasoning, rendering them useful in intricate settings and environments whereby emphasis on more relevant state transitions may be necessitated, especially in high-dimensional state spaces (Du et al., 2023; Yu et al., 2023; Radford et al., 2019). Consequently, their emergence has shown promising applications in conjunction with deep learning (Du et al., 2023; Yu et al., 2023; Kwon et al., 2023; Wu et al., 2023; Carta et al., 2023) and in particular in the domain of robotics due to their language-based instructions (Yu et al., 2023; Ahn et al., 2022).

While LLMs presents themselves as a promising basis of providing useful textual instructions (Du et al., 2023; Ahn et al., 2022), a primary limitation of these models is that their outputs can occasionally be inaccurate (Carta et al., 2023; Brown et al., 2020). Consequently, while LLMs can be useful with their language-guided prompts, it is essential to consider their shortcomings, with the most notable being that their outputs cannot be guaranteed to be optimal (Brown et al., 2020). Therefore, the implementation of LLMs calls for caution especially when directly used as either main policy drivers or as expert demonstrators for existing learning paradigms such as IL (e.g. BC and GAIL), in particular when safety-layers or mechanisms are altogether absent.

A recent work by Wu et al. (2023) utilised LLMs to extract information from game manuals via a Question Answering (QA) extraction and a reasoning module. In this process, auxiliary rewards are inferred from binary "Yes/No" answers. While an innovative approach, their work inherently requires specific instructions as well as a reasoning module to infer the auxiliary rewards. Most importantly, the main limitation of this approach is that an emergent or ideal behaviour from the agent is essentially hardwired by the content

of the manual due to the necessity to follow *specific* manual guidelines (Wu et al., 2023).

The work of Carta et al. (2023), introduced grounded LLMs. These LLMs were used to serve as the agent's policy, which was progressively updated via online RL. Nevertheless, as identified by Brown et al. (2020), the estimations and subsequent outputs from LLM cannot be guaranteed to be accurate. This, in turn, could lead to non-optimal behaviour as it is bound by the estimations stemming from the language model's output. In another work, Kwon et al. (2023), utilised LLMs to shape the behaviour of RL agents against user-specified objectives as textual prompts, to align the agent's behaviour with human-described objectives. While a novel approach, this framework was evaluated on rather short horizons compared to those commonly encountered in intricate settings and especially in sequential robotic manipulation tasks as with the works of Triantafyllidis et al. (2023b). Furthermore, their work only provided binary reward signals, without any intermediate ranges when in actuality some state transitions may still prove to be useful even though these may not be of the highest priority, so as to encourage less restrictive and more, in turn, emergent behaviour. Finally, another limitation of Kwon et al. (2023)'s work was the replacement of the traditional reward function with a proxy reward, which renders the policy quite reliant on the LLM's outputs, which is prone to inaccuracies (Carta et al., 2023; Brown et al., 2020). In a different approach, Du et al. (2023) utilised a goal-oriented approach whereby LLMs are used as the means of encouraging exploration in RL by rewarding the attainment of LLM-suggested goals based on and reliant on state descriptions to generate suitable exploration goals.

As it will be detailed further along this thesis and in particular in the final Chapter 6, the author of this thesis presents an alternative approach of utilising LLMs. In brief and in contrast to the works of Du et al. (2023); Kwon et al. (2023); Wu et al. (2023), a novel framework is presented whereby LLMs are utilised for the provision of a dense, continuous and decaying over time assistive intrinsic reward in environments naturally challenged by sparse extrinsic rewards and long-horizons. Building upon the existing RL paradigm, it is shown that this framework outperforms existing intrinsic reward methods and overcomes the limitation of utilising LLMs as main policy drivers (Carta et al., 2023; Brown et al., 2020). Moreover, the derived framework can also be integrated with existing learning methods highlighting its modularity, facilitates learning efficiency in recurring state-action (s_t, a_t) transitions and ultimately exhibits robustness against increased uncertainty levels and horizons in intricate robotic manipulation tasks.

Chapter 3

Harnessing the Sensory-Motor System via Multi-Sensory Stimulation

This chapter presents the first investigation of how different combinations of multi-sensory stimulation affect human motor performance in manipulation tasks entailing MR based technologies. From Section 2.1, it was identified that humans interact with their surrounding environment based on multiple senses, with evidence that motor movement is inherently dependent on multimodality as suggested by Billard and Kragic (2019). Naturally, this has led to an extensive research focus on multimodal interfaces and multisensory feedback as promising means of increasing overall user immersion, reflected in the ability of the sensory-motor system to coordinate meaningful motor actions based on sensory stimulation. The generations of such motor actions are invaluable during manipulation tasks in robotics, interactions in VEs and more broadly in human ergonomics (Berger et al., 2018; Billard and Kragic, 2019).

Nevertheless, an extensive comparison of the most widely used and developed sensory

Parts of this chapter are primarily based on and published as: E. Triantafyllidis, C. Mcgreavy, J. Gu and Z. Li, "Study of Multimodal Interfaces and the Improvements on Teleoperation". In: *IEEE Access*, Volume: 8, Pages: 78213-78227, 2020, doi: 10.1109/ACCESS.2020.2990080.

Some parts of this chapter are an extension of and based on: E. Triantafyllidis, "Pilot of the Future: Effectiveness of Sensory Modalities in a Multi-Modal Interface for Varying Task Complexity", M.Sc. by Research Thesis, School of Informatics, The University of Edinburgh, United Kingdom, 2019.

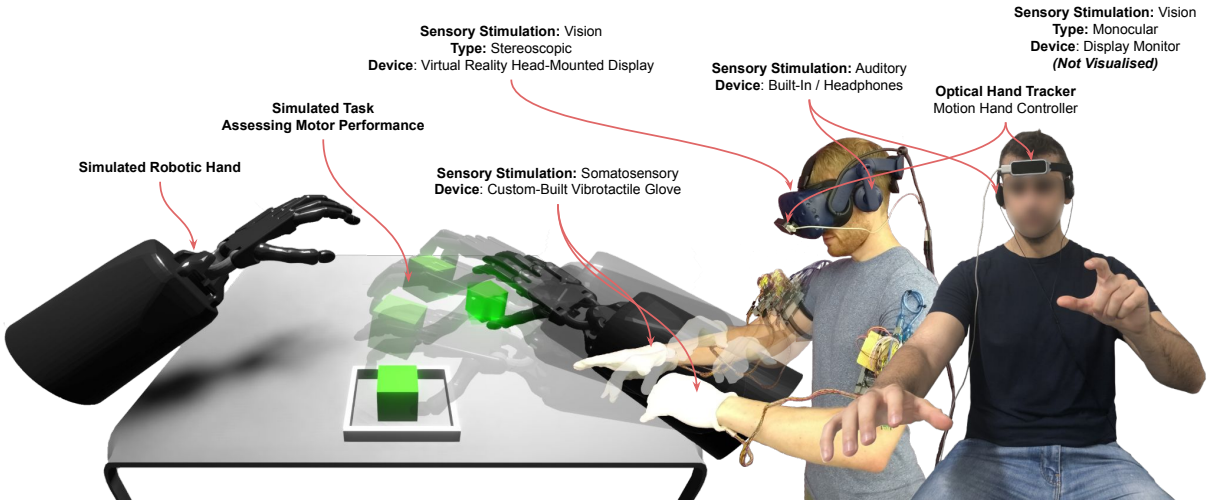

Figure 3.1: **Multimodal interfaces entailing a plethora of different sensory stimuli and their promising effects on motor performance.** The figure depicts human operators with various sensory interfaces and stimuli completing motor tasks of varying spatial complexity. This work aims to comprehensively study the effects of visual, auditory and somatosensory stimulation on the human sensory-motor system and in particular their influence on motor task performance.

stimulation of **vision**, **audition** and **somatosensory** feedback, with all their possible combinations reflected on motor manipulation tasks entailing increasing spatial complexity, remains a gap in existing work (Sigrist et al., 2013; Heilig, 1992). This chapter addresses this gap.

To this end, a comprehensive analysis with a full-factorial comparison of all possible combinations of audiovisuohaptic sensory stimulation is conducted and how these affect human motor task performance. The entailed sensory modalities of vision, auditory and somatosensory feedback accounted for a total of 2^3 (8) possible combinations. As part of the aims of this work, vision is assessed in its most widely used form of monocular vs stereoscopic viewing (Triantafyllidis and Li, 2021b). Auditory and somatosensory feedback states are assessed based on their presence or absence. Finally, motor performance is assessed on intricate, of varying difficulty motor tasks in 3D space entailing 6DoF as commonly seen in robotic teleoperation, MR interactions and simulated training scenarios (Triantafyllidis et al., 2020). Figure 3.1 visually depicts the work of this chapter. Results will show that vision is the predominant modality, exhibiting the most notable impact across all assessed parameters and measurements, emphasising the advantages of stereoscopic compared to monocular vision. Somatosensory feedback enhanced motor performance further, though not as pronounced as vision. Auditory stimulation had the

least effect on the incorporated evaluation metric, yet still exhibited marginally higher task performance when present. Overall results suggest that multi-modal interfaces prove beneficial towards human motor task performance, with tri-modal feedback surpassing bi-modal, which in turn outperforms single-modal sensory interfaces. While vision exhibited the strongest influence, incorporating other modalities should be considered due to their complementary effect on motor performance. Further investigation of the sensory design decisions governing these and other modalities is necessary to solidify and understand the sensory-motor system and how it can be harnessed via different multimodal interfaces and their role in motor performance.

Chapter Outline – This chapter is organised in a number of sections. First, Section 3.1 presents a brief introduction to the necessity of multimodal interfaces and the current gap in existing work; forming the motivation of this chapter. In Section 3.2 the formed hypotheses stemming from the literature review in Section 2.1 are presented, which are later investigated and evaluated. Section 3.3 presents the methodology, the system setup and the experimental design that allowed the construction of a multimodal interface for a comprehensive comparison of all incorporated sensory stimuli to address the gap in existing work. In Section 3.4 the different evaluation metrics are presented, the selection criteria for the human participants are outlined and the incorporated statistical analysis and procedure subsequently used for the evaluation of the data are detailed. Next, Section 3.5 presents the results stemming from the different combinations of sensory stimuli and their effect on spatially more complex motor tasks. Finally, in Section 3.6 the results are discussed and framed with the findings of existing work. Chapter 7 summarises the main conclusions and future work of this chapter.

3.1 Introduction and Problem Statement

With the spike in MR technologies in recent years, simulation environments entailing high-fidelity and realistic physics are becoming increasingly more immersive. This, in turn, has led to the elucidation of human perception, leading to promising directions towards maximising human performance (Barrera Machuca and Stuerzlinger, 2019; Triantafyllidis and Li, 2021a). Simulating and incorporating such technologies in telepresence and especially teleoperation scenarios have seen a promising rise in operator perceptive skills corresponding to increased motor task performance and more broadly in human

ergonomics (Triantafyllidis et al., 2020; Triantafyllidis and Li, 2021b).

In the context of teleoperation, a human operator controls a remote robotic system whereby, usually, a set of motor tasks are delegated. In this process, the instinct, complex cognition and high motor dexterity unique to humans with the computational and physical capabilities of robots can be combined. Unfortunately, during telepresence and especially teleoperation cases, the actions between the human and the remote robotic system are physically detached from one another, which constitutes the overall experience unnatural as argued by Triantafyllidis et al. (2020). This limitation appears to also hold true during the more broad context of avatar embodiment in virtual environments when concerned with manipulation (Triantafyllidis et al., 2020). In particular, this fusion of biology and machine inherently relies on an interface that can stimulate the human senses in such a way that the virtual embodiment is, in theory, indistinguishable from real life, or from a practical standpoint, close to "ideal" (Sigrist et al., 2013; Chen et al., 2007). To achieve this, one crucial aspect should be prioritised, namely – the human sensory-motor system and its workings.

Current research in multimodal interfaces aims to provide solutions to increase the ability to elucidate the human senses and more broadly the perceptive system, to generate motor movements and actions with higher efficiency and performance (Yanco and Drury, 2004; DeJong et al., 2004; Jennett et al., 2008).

From a biological perspective, humans, in the general context of perception, utilise a wide range of stimuli to perceive their surroundings including visual, auditory, somatosensory, olfactory and gustatory feedback (Heilig, 1992; Billard and Kragic, 2019). From Section 2.1, it was identified that in the related work, it is generally argued that manipulation and the general interaction of humans with their environment are inherently multimodal, necessitating the simultaneous activation of a plethora of senses to successfully meet environment demands (Billard and Kragic, 2019). Related neuroscience studies findings support that the simultaneous presence of two or more sensory stimuli, in particular visual and somatosensory sensory cues, are beneficial particularly due to both modalities overlapping in the same brain region (A Ghazanfar and E Schroeder, 2006; Amedi et al., 2001; James et al., 2002; Sathian and Zangaladze, 2002). Narrowing the interaction of humans with their environment to the context of manipulation and motor movement, the primary stimuli that are in use and relevant include visual, auditory and somatosensory feedback (Sigrist et al., 2013; Triantafyllidis et al., 2020). These stimuli are also supported

in being more developed than the gustatory and olfactory senses according to Popescu et al. (2002). Heilig (1992) estimated that vision, audition, and somatosensory stimuli have different sensory contributions, amounting to 70%, 20% and 4% respectively.

However, from the related work and to the best of the author's knowledge, it was inferred that an extensive comparison of how the different sensory combinations of **vision**, **audition** and **somatosensory** feedback quantitatively contribute towards motor performance of increasing complexity is missing according to Sigrist et al. (2013). This chapter aims to address this gap. Simulating *olfactory* as well as *gustatory* stimuli were disregarded due to (i) their limited relevance from a motor movement perspective (Triantafyllidis et al., 2020), (ii) their limited sensory contribution to other senses (Popescu et al., 2002; Heilig, 1992) as well as (iii) the complexities and resources associated with building an appropriate apparatus (Heilig, 1992). The work outlined in this chapter address *this* gap in the literature and provides a novel comprehensive comparison of audio-visuo-haptic sensory modalities and their contribution towards motor performance across tasks of varying spatial complexity. Figure 3.1 visually depicts the work carried out in this chapter.

3.2 Hypotheses

From the literature review in Section 2.1, a number of hypotheses were formed to adequately study how audiovisuohaptic interfaces, and by extent of all their possible combinations, quantitatively affect human motor performance via subjective and objective metrics. The hypotheses formed are summarised as follows:

Hypothesis H1: *Lower cognitive workload will be perceived and by extent corresponding to higher motor performance with (a) the use of a stereoscopic VRHMD than with a generic monocular display monitor, (b) the presence of somatosensory feedback compared to its absence, and finally (c) the presence of auditory feedback than its absence.*

Hypothesis H2: *Higher perceived system usability corresponding to higher motor performance with (a) the use of a stereoscopic VRHMD than with a generic monocular display monitor, (b) the presence of somatosensory feedback rather than its absence, and finally (c) the presence of auditory feedback than its absence.*

Hypothesis H3: *Faster motor performance and by extent corresponding to less task completion time will be observed with (a) the use of a stereoscopic VRHMD than with a generic monocular display monitor, (b) the presence of somatosensory feedback rather than its absence, and finally (c) presence of auditory feedback than its absence.*

Hypothesis H4: *Improved depth estimation corresponding to lesser spatial error will be measured via (a) the use of a stereoscopic VRHMD than with a generic monocular display monitor, (b) the presence of somatosensory feedback than without, and finally (c) the presence of auditory feedback than without.*

Hypothesis H5: *Higher spatial accuracy will be measured via (a) the use of a stereoscopic VRHMD than a generic monocular display monitor, (b) the presence of somatosensory feedback than its absence, and finally (c) the presence of auditory feedback than without.*

3.3 Methodology

This section presents the system apparatus, including the system architecture, incorporated hardware and software components to accommodate the study and test the formed hypotheses to address the existing gap in the literature. Foremost, a series of experiments were designed, during which the participants performed a pick-and-place task under various conditions of increasing motor complexity. A full factorial comparison was conducted in this experiment whereby all possible combinations of a visual *(monocular vs stereoscopic)*, auditory *(presence vs absence)* and somatosensory *(presence vs absence)* interface were assessed, with more details in Table 3.1 and visually shown in Figure 3.1.

3.3.1 System Apparatus

A notable amount of time was dedicated to setting up the simulation environment and accommodating all the necessary software plugins and hardware used in the physics-enabled game engine Unity3D, acting hereinafter as the primary simulation engine facilitating. The particular system apparatus was subject to minimal changes and iterations further along the PhD project (particularly for Chapter 4), constituting it in large the basis for running subsequent simulations and experiments.

Table 3.1: **Full factorial comparison of the audiovisuohaptic interface.** The multimodal interfaces are broken down into the 2^3 (8) possible combinations of visual, auditory and haptic feedback states.

	Vision		**Audition**		**Haptics**	
	Monitor	VRHMD	Absence	Presence	Absence	Presence
Condition 1	X		X		X	
Condition 2	X		X			X
Condition 3	X			X	X	
Condition 4	X			X		X
Condition 5		X	X		X	
Condition 6		X	X			X
Condition 7		X		X	X	
Condition 8		X		X		X

To ensure consistent frame rates and most importantly physics performance in the simulation engine, a resource-intensive Graphics Processing Unit (GPU) was of necessity. Consequently, an NVIDIA 2080 RTX Ti GPU was used as the main GPU, furthermore ensuring stable frame rates for the VRHMD later on used for the experiments.

The primary tool used to interact with the simulation was the Leap Motion Hand Controller (LMHC), an optical hand tracking device equipped with two infrared cameras providing stereo vision at 120Hz and a 135-degree FoV. With the LMHC, the ability to manipulate objects in the simulation was rendered feasible. To enhance the visual experience and conduct a comparison of the visual sensory feedback, a generic monocular display monitor and a VRHMD were used. Additionally, a custom data glove was developed that was capable of stimulating vibrotactile feedback for somatosensory stimulation, which is described in the following section. Figure 3.2 provides an overall overview of the simulation setup.

3.3.2 Hand Manipulation and Control

From the LMHC, the Cartesian joints positions in the world frame were retrieved via the accompanied Leap Motion Software Development Kit (SDK). Those joint positions were then translated to joint angles which were subsequently used to directly control the virtual hand in the simulation. This translation was made possible by calculating the

angle θ between a child joint $\vec{b_{i-1}}$ and its parent joint $\vec{b_i}$:

$$\theta = \arccos\left(\frac{\vec{b_i} \cdot \vec{b_{i-1}}}{\left\|\vec{b_i}\right\| \left\|\vec{b_{i-1}}\right\|}\right). \tag{3.1}$$

The implementation of a Proportional Derivative (PD) controller was used to control the joints. Each individual joint of the anthropomorphic Shadow dexterous hand used in the simulation was controlled by an individual PD controller, which was formulated as follows for each time-step t:

$$u(t) = K_P \cdot e(t) + K_D \cdot \dot{e}(t), \tag{3.2}$$

where $u(t)$ is the angular velocity control signal sent to each individual joint. The term $e(t)$ is formulated as $e(t) = q_h(t) - q_r(t)$, representing the current position error between the human joint and the robot joint and $\dot{e(t)}$ formulated as $\dot{e(t)} = \dot{q}_{ref} - \dot{q}_r(t)$, representing in turn the velocity error between the robot and the desired velocity, which in this particular case is set to zero. The K_P and K_D terms represent the gains which were tuned accordingly such that both the human and robot motions were matched as accurately and as smoothly as possible. The hand manipulation approach implemented and described above is visually depicted in Figure 3.2.A.2.

3.3.3 Design of Multimodal Interfaces and Sensory Feedback

Next, the multimodal interfaces were designed, entailing the necessary hardware later used in the experiments. As mentioned further in this chapter, acclimatisation to all the controls, technologies and hardware was allowed for the participants prior to the commencement of the experiments.

Design of Visual Stimulation

For visual feedback, two states were compared; monoscopic visual feedback from a generic 2D display monitor and stereoscopic vision stemming from a VRHMD. These represented the most used display hardware in related work concerned with the manipulation of objects during robotic teleoperation and simulated environment training (Triantafyllidis and Li, 2021b). In addition to the visual disparity between these devices, the VRHMD also allowed users to control their viewpoint in the virtual environment by moving their heads. To this end, to conduct a fair experiment, participants were also able to change their viewpoint within the simulation when using the monitor by using a generic computer keyboard using a set of predefined keybindings. To retain consistency, the optical hand

Figure 3.2: **Simulation, hand-control and interface overview.** The figure depicts the overall apparatus, including the simulation overview, the employed hand-control procedure and snapshots of the simulation. **Figure A.1:** The development of the custom-built somatosensory capable haptic glove. Early and final development stages are depicted. **Figure A.2:** The employed hand-retargeting approach, tracking the hand of the human participants in the real world (either with or without the glove), the corresponding tracked hand in the simulation and the final hand-retargeting to the robotic hand. **Figure B.1:** The simulation overview with all software used, arrows indicate information and data flow. **Figure B.2:** The overview of the motor task in the simulation with the hand-re-targeting approach in full working order. **Figure C:** Participant recordings in the simulation performing the different motor task experiments across varying sensory interfaces.

controller LMHC, was in a head-mounted state for both the VRHMD and the display monitor. For the latter case, a headband was used to retain a head-mounted state for the LMHC while for the stereoscopic case, the optical tracker was attached at the front of the VRHMD. Figure 3.1 depicts the physical design of the visual modality.

In terms of their hardware specifications, a monitor measuring 27-inches for monocular viewing was utilised, with an In-Plane Switching (IPS) display of 2,560 x 1,440 resolution and a 60Hz refresh rate placed 75-100cm away from the participants. The VRHMD used was the HTC Vive Pro with a 3.5 inch Active-Matrix Organic Light-Emitting Diode (AMOLED) screen at 2,880 x 1,600 resolution (1,440 x 1,600 pixels per eye), 90Hz refresh rate and 110° FoV. High-resolution displays were chosen to limit distance overestimation and degraded longitudinal control which are common issues in virtual environments as supported by B F van Erp and Padmos (2004).

Design of Auditory Stimulation

Auditory feedback was deemed important due to its potential benefit as an alternative supplementary sensory pathway; compared to solely depending on the superimposition of visual information (Sigrist et al., 2013). For auditory stimulation, commercial off-the-shelf hardware was used. For the stereoscopic visual case VRHMD a built-in audio headset was available while for the monocular case, an external generic audio headset was provided. The audio quality was set at 16-bit and 44,100Hz. The auditory case was designed in such a way as to resemble everyday sound effects such as scraping or collisions of objects within the simulation environment. This, in turn, minimised the necessity for prior context, similarly to Triantafyllidis et al. (2020). Furthermore, terminal and concurrent auditory feedback were also presented as the means of indicating task-specific states in the form of alarms. These corresponded to (i) environment collisions, (ii) task time indicators (iii) task success sounds. Lastly, picking up, dropping or placing the objects produced realistic bump and scrape sounds that one would expect when interacting with real objects. Figure 3.1 depicts the physical design of the auditory modality.

Design of Somatosensory Stimulation

Finally, for somatosensory feedback, a custom-built, low-cost haptic vibrotactile-capable glove was designed and constructed. The decision to create a haptic glove stemmed from the fact that preliminary findings indicated that vibrotactile stimulation when pre-

sented with an appropriate synchronisation with other modalities, is sufficient for realising grasped objects and a promising alternative to force feedback (Berger et al., 2018; Weber et al., 2013; Aleotti et al., 2002; Suzuki and Kobayashi, 2005) and in some cases more beneficial in direct manipulation tasks (Kontarinis and Howe, 1995; Massimino and Sheridan, 1993). Figure 3.1 and Figure 3.2.A.1 depict the physical design of the somatosensory modality.

The construction of the haptic glove was greatly inspired by the work of Kim et al. (2017). More specifically, to provide haptic sensation, a set of 15 coin-vibration Direct Current (DC) 3V 70mA 12,000 Revolutions Per Minute (RPM) motors were used, two of which were placed on each proximal & distal phalanges of the glove, and five motors in total to cover the larger area of the palm. Wireless communication between the virtual environment and the glove ensured free movement and was made possible via a Bluetooth transceiver connected to an Arduino microcontroller acting as the main driver for each glove. The early and final development stages of the haptic glove are shown in Figure 3.2.A.

The vibration intensity for each of the motors was proportional to the force induced by the virtual object the user was in collision with. More specifically, physical properties were incorporated including kinetic energy and object penetration for the manipulation experiments to induce vibrotactile stimulation. In particular, vibration intensity was proportional to the kinetic energy K_E and to the object penetration P, with the latter allowed in the simulation. These two elements combined provided the final intensity.

Foremost, the kinetic energy K_E of the virtual collision was formulated as follows:

$$K_E = \frac{1}{2} \cdot m \cdot v^2, \tag{3.3}$$

whereby m is the mass of the body and v is the velocity between the robot segment and the environment.

Secondly, for the virtual penetration between the simulated robotic hand and the objects within the simulation, the relative penetration of P between the robot and the environment was used as a proxy for force. Penetration is defined by the relative distance between the robot segment v_r and virtual object v_o and the distance between the centre and surface of the object s_o as shown in Equation 3.4,

$$P = \left| 1 - \frac{\|v_r - v_o\|}{\frac{1}{2} \cdot s_o} \right|. \tag{3.4}$$

Consequently Equation 3.3 and Equation 3.4 were combined to calculate the final normalised vibration intensity shown in Equation 3.5,

$$V = V_{min} + a \cdot \frac{(V_{max} - V_{min})}{K_{E_{max}}} \cdot K_E + b \cdot \frac{(V_{max} - V_{min})}{P_{max}} \cdot \ P, \tag{3.5}$$

whereby V is the final vibration intensity transmitted to the vibration motors, V_{min} is the minimum vibration intensity needed to distinguish vibrotactile stimulation when in contact. V_{min} was set to 25% based on a pilot study consisting of five participants to allow minimal vibrotactile feedback during collisions. The second term calculates the vibration intensity based on the kinetic energy exerted and is controlled by a constant a. V_{max} is the maximum vibration intensity of the hardware. $K_{E_{max}}$ is the maximum calculated kinetic energy in Joules with a velocity limit of $v = 7\text{m/s}$ set in the physics engine (PhysX) within Unity3D and K_E is the current kinetic energy exerted to the virtual object. The final term of Equation 3.5 represents the vibration intensity based upon the penetration of the robotic hand with the virtual object and is controlled by a constant b. P_{max} is the maximum penetration allowed. Finally, the term P is defined as the current penetration exerted on the object. Figure 3.2.A.1 visually depicts the custom-built haptic glove with its electronics, drive control board and motors exposed.

It is worthwhile to note that the haptic glove did not include any Inertial Measurement Unit (IMU) for tracking its position. Instead, the LMHC tracked the haptic gloves at all times, as anthropomorphic features were retained, as shown in Figure 3.1 and Figure 3.2.A.

3.3.4 Design of Tasks with Increasing Motor Complexity

All tasks in the experiments conducted required the participants to pick up 3D primitives from a set starting position and to be placed in a randomised target location within a predefined time limit. The primitive 3D objects chosen for this study were a (i) Cube, (ii) Cylinder and a (iii) Sphere. The decision to choose such primitive shapes was twofold; (i) these represent the most used 3D objects for virtual training studies and (ii) their different shapes allow for the assessment of different grasping techniques (Kumar and Todorov, 2015; Light et al., 2002).

Light et al. (2002) introduced a clinically validated hand function test, the SHAP. The incorporated three primitive shapes in this work encompass most of the six distinctive grasping classifications of SHAP, which include: **precision-tip**, **lateral**, **tripod**, **spherical**, **power** in addition to **extension** grasping. Figure 3.3.B illustrates the SHAP grasping

Figure 3.3: **The overview of the motor tasks of varying spatial complexity including the different grasping classifications. Figure A:** The figure depicts the simulation overview of the experimental tasks of varying spatial complexity utilised for the evaluation of the effectiveness of the sensory interfaces studied. **Figure A.1:** The manipulation motor tasks illustrating the three different primitive 3D shapes, their sizes as well as distances from start to target. Three distances from start to target were evaluated; 15cm, 30cm and 60cm, highlighted in colour for visualisation purposes in green, yellow and red respectively. **Figure A.2:** The figure depicts all 18 tasks studied, broken down into (i) three object shapes, (ii) two object sizes and (iii) three spatial distances. Change in lettering, A, B, or C indicates different shape types. Change in the first digit, A.1, A.2, B.1, B.2, C.1 and C.2 indicates different shape sizes within the shape type category; with values 1 and 2 corresponding to shape sizes 5cm and 3cm respectively. Change in the last digit, A.1.1, A.1.2 A.1.3 etc., indicates increased spatial distance within sub-scale shape size and shape type. In particular, for notations A.1.1, A.1.2 A.1.3, the last digit of values 1, 2 and 3 corresponds to 15cm, 30cm and 60cm distance from start to target placement respectively. For more information on the notation of the 18 tasks (i.e. A.1.1 to C.2.3), consult Section 3.3.4. **Figure B:** The six distinctive grasping classifications based on the SHAP test in the simulation with the Shadow dexterous hand. Participants were allowed to use any of the six grasping classifications. **Figure B.1 to B.6:** From left to right, (B.1) Tip, (B.2) Lateral, (B.3) Tripod, (B.4) Spherical, (B.5) Power and (B.6) Extension grasping classification.

classifications on the anthropomorphic robotic hand in the simulation environment.

More specifically, three in total different tasks based on the aforementioned three primitive 3D shapes were derived, in particular:

- *Task I – Cube Manipulation:* For the first task, a generic 3D cube was manipulated. A cube shape was used as its inherent spatial complexity is lower than the other two shapes as a cube does not flip or roll, rendering its difficulty lower. Grabbing techniques employed for this shape included Precision Grasping via Palmar Pinch (Light et al., 2002).
- *Task II – Cylinder Manipulation:* The second task included the manipulation of a cylinder shape and unlike a cube, a cylinder can flip over and roll over a surface, rendering the task inherently harder. The grabbing techniques employed here, according to the SHAP, included Precision Grasping via Palmar Pinch, as well as Cylindrical Grasping, also known as Power Grasping (Light et al., 2002).
- *Task III – Sphere Manipulation:* The third and final task included the manipulation of a sphere. The sphere, unlike the cube or the cylinder, was considered to be the most complex task due to its inherent ability to roll over a surface. Grabbing techniques employed included Precision Grasping via Palmar Pinch as well as Spherical Grasping (Light et al., 2002).

Motor complexity in the context of this work is defined as the spatial complexity, or layout, of a given task. In particular, the following are presented in ascending order of motor task complexity: (i) the Cube, Cylinder and Sphere shapes, (ii) the smaller dimensions of these and (iii) the increasing distances between the starting and target position. In terms of spatial positioning, the distance from the set starting position to a random target location ranged from 15.0, 30.0 and 60.0 (cm), while the sizes of shapes were classified as either large 5.0 x 5.0 x 5.0 (cm) (LxWxH) or small 3.0 x 3.0 x 3.0 (cm) (LxWxH). The spatial complexity of the aforementioned is visually represented in the form of a tree in Figure 3.3.A.2.

A total of 144 trials were conducted per participant taking part in the study, stemming from the three different types of 3D primitive shapes, three different distances, two different sizes and across eight in total multimodal interface conditions (3x3x2x8). With a total of $N = 25$ participants, a grand total of 3,600 trials were recorded. The motor tasks of varying spatial complexity are visually depicted in Figure 3.3.A.

Tasks were either classified as a success or as an error before proceeding to the next. Success was defined when the object was placed at the designated random location, regardless of the spatial accuracy even though this was assessed later on in the analysis. Each task had a maximum completion time set at $t = 30s$, after which the task ended. If by the end of this predefined time, the object was not within the bounds of the target location, the was considered an error. For all tasks, an invisible collision wall was implemented to avoid objects falling out of physical bounds within the task space set.

3.4 Experiments and Evaluation

This section presents how the experiments were conducted and the subsequent evaluation, as part of the data acquired from them. A number of different evaluation methods were utilised to assess the effectiveness of the sensory modalities reflected on increasingly more complex motor tasks.

3.4.1 Subjective Measurements

Foremost, to assess the subjective responses by the participants, two questionnaires were implemented: (i) the System Usability Scale (SUS) assessing system usability proposed by Brooke et al. (1996), and the National Aeronautics and Space Administration - Task Load Index (NASA-TLX) assessing workload demand proposed by Hart and Staveland (1988). Both of these questionnaires consisted of Likert scale responses allowing the quantitative evaluation of each participant's subjective feeling after each sensory interface condition, as outlined in table Table 3.2.

3.4.2 Objective Measurements

For objective measurements, a series of time-based and spatial-based metrics were incorporated. Each of these measurements was collected at the end of each sub-task. Task successions as a means of percentage completion were calculated. For time-based metrics, placement and completion times were measured while for spatial metrics, target error including position and rotation accuracy to the target placement were recorded.

Placement times were defined as the duration it took participants to pick up the object and place it at the target location. Contrarily, completion times were defined as the overall time it took participants to complete the entire task successfully while retaining

Table 3.2: **Incorporated evaluation metrics to assess the effectiveness of different sensory interfaces and stimuli on motor task performance across a variety of spatial complexity tasks.** Summary of both the objective and subjective evaluation measurements.

Measurement	**Type**	**Metric**
Cognitive Workload (NASA-TLX)	Subjective	Questionnaire [*Likert Scale*]
System Usability (SUS)	Subjective	Questionnaire [*Likert Scale*]
Task Succession	Objective	Percentage [%]
Placement Time	Objective	Seconds [s]
Completion Time	Objective	Seconds [s]
Target Error	Objective	Meters [m]
Position Accuracy	Objective	Percentage [%]
Rotation Accuracy	Objective	Percentage [%]

the object. This decision was made to analyse the time it strictly took the participants to move the object to the target volume (*placement time*) and thereafter the additional time for any accuracy and spatial corrections (*completion time*). In this way, both motor and spatial performance could be evaluated.

All subjective and objective measurements incorporated are shown in Table 3.2.

3.4.3 Participants

For this study, a total of ($N = 25$) participants were recruited via an advertisement at the University of Edinburgh. The ages of subjects ranged from 21 to 44 ($M = 26.36$, $SD = 4.829$, $Mdn = 26$), of which 6 were females and 19 males. Each participant had healthy hand control, normal hearing ability and normal or normal-to-corrected vision. A 30-minute interactive experience using the VRHMD was given as compensation at the end of the experiments.

Prior to the commencement of the experiments, participants were briefed on the purpose of the experiment and the aims of the study, gave formal written consent and were handed out the NASA-TLX as well as the SUS questionnaires to allow acquaintance with the scales. Once the participants familiarised themselves with the questionnaires, their Inter-Pupillary Distance (IPD), was measured for the VRHMD and they were allowed 10

minutes to familiarise themselves with the simulation environment. During this acclimatisation procedure, participants conducted a set of training exercises, entailing using the keyboard controls, the VRHMD as well as the haptic glove. These training exercises were specific to the acclimatisation procedure and independent of those in the experiment to avoid familiarity with the actual evaluation tasks. Due to having eight different sensory interface conditions, the order of these multimodal interfaces was further randomised for each participant to minimise task adaptation.

3.4.4 Statistical Analysis and Procedure

For all subsequent results presented, a significance level of 95% CI was set. Foremost, a ANalysis Of VAriance (ANOVA) and more specifically a Repeated Measures - ANalysis Of VAriance (RM-ANOVA), to analyse the parametric results that did not violate normality via a Shapiro-Wilk Test. Moreover, a post-hoc analysis for the pairwise comparison of the eight different interface conditions was implemented to allow for an increased understanding of all possible combinations of those three tested sensory modalities. Thereinafter, Maulchy's Test of sphericity was used to test sphericity and in cases of violation, a Greenhouse–Geisser Correction (GGC) was used to account for the violation and correct the degrees of freedom assuming an of $\epsilon < .75$. Data violating normality via the Shapiro-Wilk Test i.e. non-parametric data, in particular, ordinal ones i.e. Likert scales, an Aligned Rank Transform (ART), proposed by Wobbrock et al. (2011a), was used to allow the use of parametric tests such as the RM-ANOVA on the ranked data. For non-parametric continuous data on the other hand, a Friedman's test, similar to the RM-ANOVA, was used to test for significance across the eight multimodal interface conditions (Friedman, 1937). For post-hoc analysis, a Wilcoxon signed-rank test was performed for the subsequent pairwise comparison across the interface conditions. Additionally, for samples that were classified as a Bernoulli distribution, the proportion of successful completion, a two times standard deviation (2σ) from the mean were considered significant (95% CI), as per the empirical rule (Pukelsheim, 1994).

Hereinafter, the significance levels are denoted as: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$ and n.s (or $p > .05$) standing for "not significant". Finally, the overall results of each multimodal interface condition across all measurements are summarised, thus providing new evidence to the hypothesised and untested effectiveness of each interface condition suggested by Sigrist et al. (2013), while reproducing a similar to their work visual representation.

3.5 Results

In this section, the results are presented from the experiments conducted with all eight multimodal interface conditions, which are thereafter comprehensively analysed. This analysis subsequently aided in a comprehensive understanding of the sensory-motor system and in particular the influencing sensory modalities on human motor performance, commonly seen in robotics and MR interactions in full 3D space entailing 6DoF (Triantafyllidis and Li, 2021a; Triantafyllidis et al., 2020).

3.5.1 Subjective Results

All of the subjective results conducted entailed questionnaires. Since these questionnaires were classified as ordinal data following Likert scales, an ART was applied prior to conducting an RM-ANOVA to allow the analysis of the ranked data.

Perceived Workload - NASA-TLX

Foremost from the results, a highly significant effect on the perceived workload (NASA-TLX) across all interface conditions was observed via a one-way RM-ANOVA yielding ($F(2.393, 57.437) = 70.473$,$p < .001$,$\eta_p^2 = .746$), with a GGC applied ($\epsilon = .342$).

A subsequent post-hoc analysis revealed partial support of *Hypothesis H1.* More specifically, conditions incorporating the use of monocular vision via a generic display monitor (C1, C2, C3 and C4) accounted for significantly higher perceived workload ($p < .001$) than stereoscopic vision with the use of a VRHMD (C5, C6, C7 and C8) (*Hypothesis H1.a*). Additionally, those multimodal interface conditions incorporating stereoscopic vision with the addition of somatosensory feedback (C6 and C8), showed further reduced perceived workload ($p < .05$) than those that do not (C5 and C7). The difference between interface conditions with monocular visual feedback coupled with and without somatosensory sensation (C2 and C1 respectively), was marginally lower, but not to a significant degree ($p = .056$). Thus *Hypothesis H1.b* was partially supported. Finally, *Hypothesis H1.c*, was not supported as all conditions with auditory feedback did not contribute to an observable difference in perceived workload ($p > .05$). The NASA-TLX results are shown in Figure 3.4.A and Table 3.3.

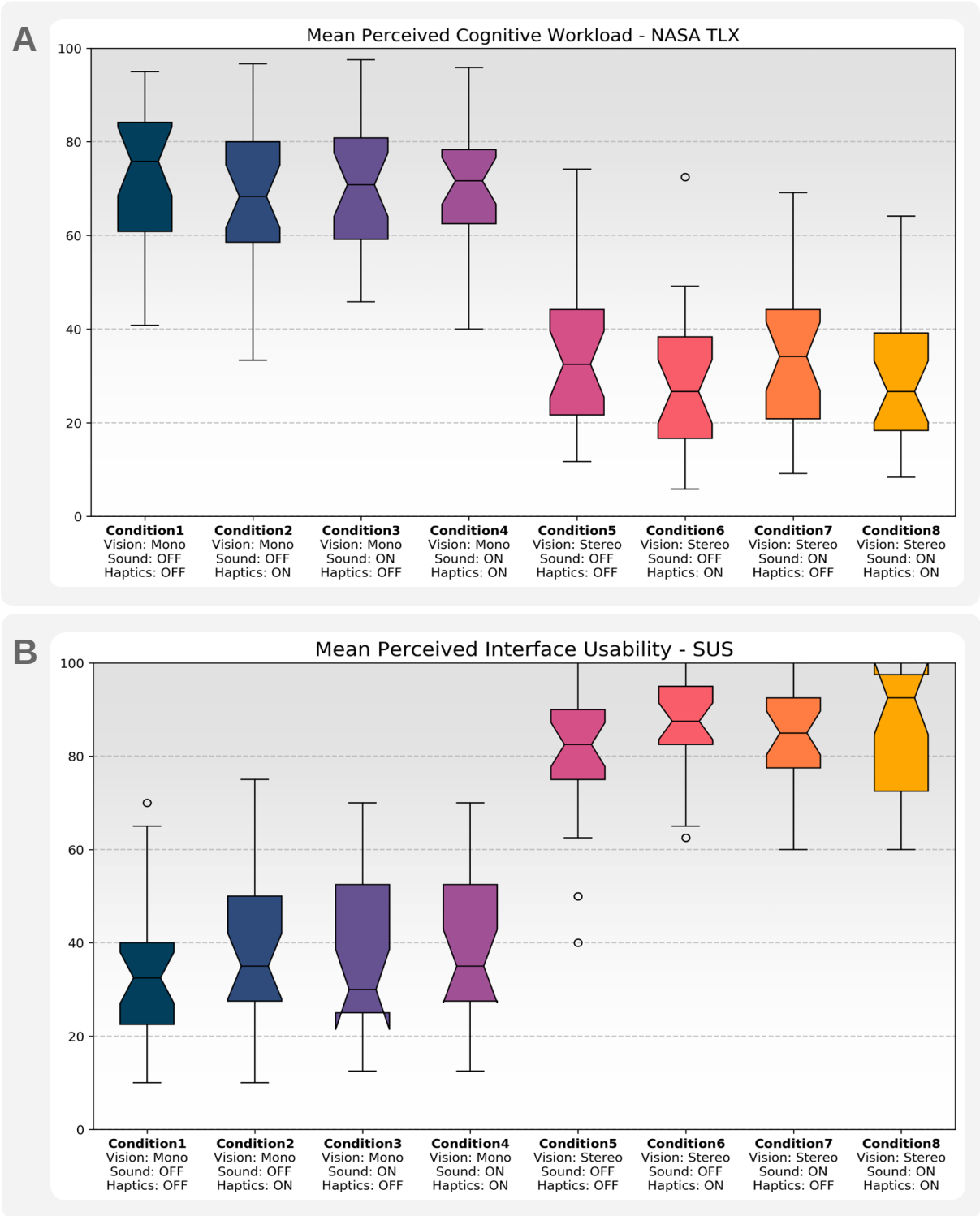

Figure 3.4: **Subjective results for sensory effectiveness on motor task performance stemming from the SUS and NASA-TLX responses.** Box plots results stem from $N = 25$ participants. Dots represent outliers. **Figure A:** The NASA-TLX box plot illustration of the mean perceived mental workload, with higher scoring equaling worse performance. **Figure B:** The SUS box plot illustration of the mean perceived interface usability, with higher scoring equaling to better performance.

Table 3.3: **The summary of all the subjective results on the perceived motor performance based on different sensory interfaces, reporting median and standard deviation across all eight interface conditions.** Higher values for the NASA-TLX correspond to higher cognitive demand and by extent to worse performance, while higher values for the SUS correspond to higher interface usability i.e. increased performance.

Subjective Metrics				**NASA-TLX**		**SUS**	
	Vision	*Audio*	*Haptic*	*Med.*	*Std. D.*	*Med.*	*Std. D.*
C1	Monitor	Off	Off	75.83	±13.82	32.50	±14.34
C2	Monitor	Off	On	68.33	±16.22	35.00	±17.15
C3	Monitor	On	Off	70.83	±15.00	30.00	±17.76
C4	Monitor	On	On	71.66	±14.93	35.00	±15.63
C5	VRHMD	Off	Off	32.50	±16.73	82.50	±15.82
C6	VRHMD	Off	On	26.66	±15.91	87.50	±11.33
C7	VRHMD	On	Off	34.16	±15.58	85.00	±11.38
C8	VRHMD	On	On	26.66	±14.52	92.50	±13.20

Perceived System Usability - SUS

Similarly to the perceived workload, an ART was performed prior to the use of the RM-ANOVA on the ordinal data. In line with the results from the perceived workload, a highly significant effect on the perceived system usability (SUS) across all interface conditions was observed, as yielded by a one-way RM-ANOVA ($F(3.930, 94.310) = 97.064$,$p < .001$,$\eta_p^2 = .802$), with a GGC ($\epsilon = .561$).

Post-hoc analysis thereafter revealed the same trend to hold true for the system usability as with that of the NASA-TLX analysis. In particular, similar partial support of *Hypothesis H2.a* was observed, as stereoscopic vision via the use of the VRHMD (C5, C6, C7 and C8) accounted for significantly higher interface usability, ($p < .001$), than merely using a generic display monitor with monocular vision (C1, C2, C3 and C4). Support of *Hypothesis H2.b* was again partial, as somatosensory feedback further increased the overall system usability limited to cases only via the use of stereoscopic visual feedback (C6 and C8), ($p < .05$). Finally, auditory feedback did not yield any significant observable difference in perceived system usability across all interface conditions ($p > .05$). The SUS results are shown in Figure 3.4.B and Table 3.3.

3.5.2 Objective Results

For the objective results, a mixture of time and spatial-based metrics were recorded to evaluate the performance of the participant completing the motor tasks in the virtual environment. For details regarding these measurements consult Table 3.2. Similarly to the subjective results, data that violated normality via a Shapiro-Wilk Test, was adjusted via an ART prior to the use of the RM-ANOVA to allow the analysis of the ranked data. Alternatively, a Friedman's test was utilised, which did not require the use of the ART.

Error-Based Evaluation - Completion of Tasks

Foremost, the proportion of successfully completing a given task as a percentage rate (%) was evaluated. This was recorded for every task across all eight interface conditions. The sample was analysed and was classified as a Bernoulli distribution. For the subsequent evaluation of determining the significance of the data, a two-times standard deviation from the mean, also known as the three-sigma or empirical rule, was used to test for the significance of the results.

From the results and as similarly observed with the NASA-TLX and the SUS questionnaires for the subjective results, sensory interface conditions incorporating a VRHMD (C5, C6, C7 and C8), accounted for significantly higher mean success rates ($p < .05$) than those with visual feedback stemming from a monocular display monitor (C1, C2, C3 and C4). More specifically, the success rates for interface conditions with a monocular display were (C1 to C4) 39.33% (SD=21.69%), 47.55% (SD=18.77%), 51.33% (SD=23.63%), 48.22% (SD=20.26%) respectively. Contrarily, success rates of the conditions utilising the VRHMD were (C5 to C8) 96.22% (SD=4.73%), 99.11% (SD=2.62%), 96.22% (SD=5.94%), 97.55% (SD=4.26%) respectively. In contrast to the subjective results, the error-based evaluation showed no significant observable difference between conditions that incorporated the presence or absence of either haptic or auditory feedback, which were $p > .05$ in both cases. These results are depicted as a heat map visually in Figure 3.5.A.

Time-Based Evaluation - Placement and Completion Times

In contrast to the aforementioned incorporated metrics, for the time-based metrics only data that stemmed from successful instances were considered. This, in turn, transformed the data into a non-parametric state as shown by a Shapiro-Wilk Test showing a significant deviation from normality for both placement and completion times ($p < .05$). Conse-

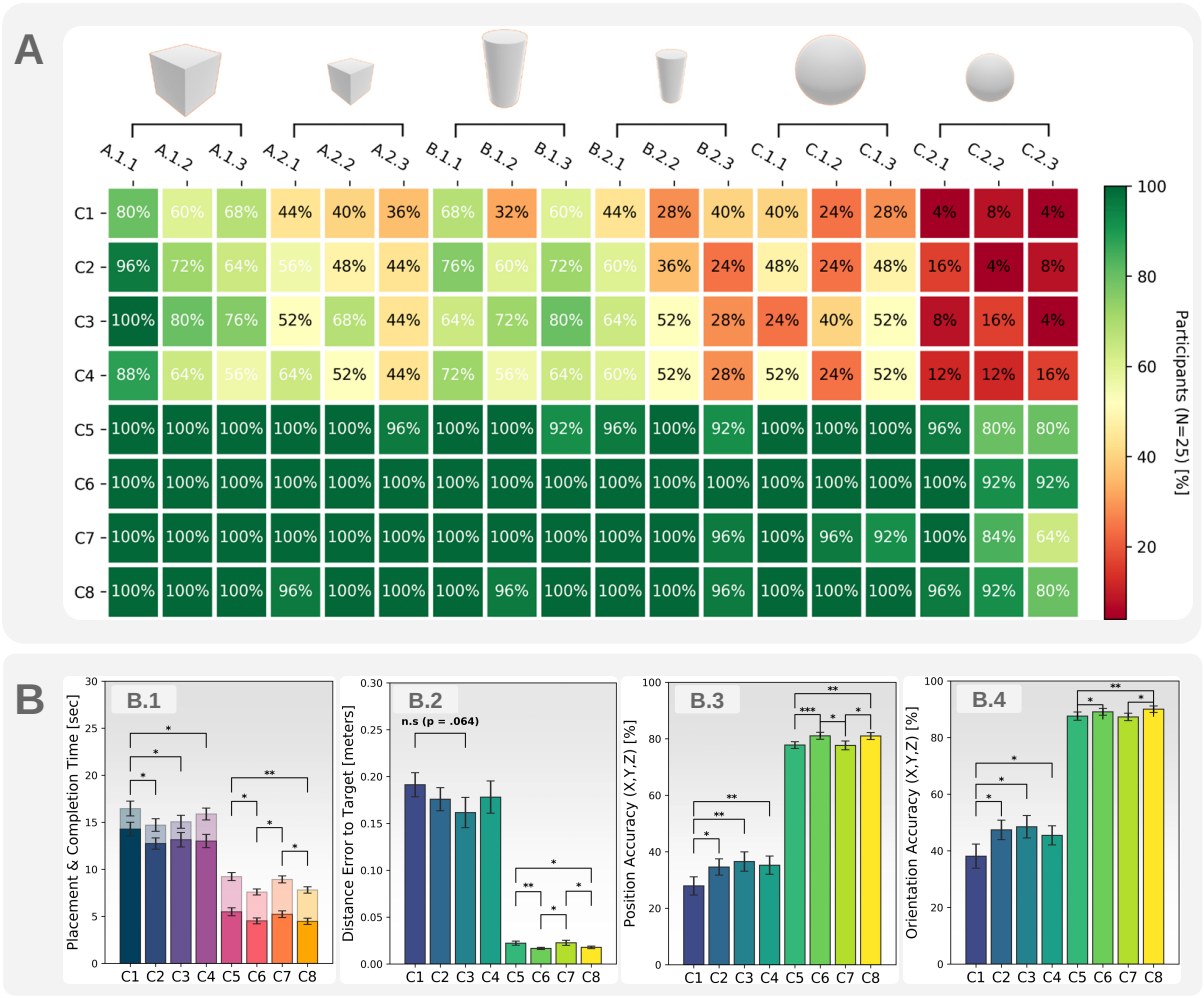

Figure 3.5: **Objective results for sensory effectiveness on motor task performance entailing error rates, time and spatial-based metrics. Figure A:** Heat-map illustration of the proportion of task success rates and errors. Increased spatial complexity from left to right; including increased spatial distances, shape sizes and shape types. Spatial difficulty is overall ascending from horizontal axis A.1.1 (left) to C.2.3 (right). Task success is indicated as percentages, with colouring from red to green indicating improved performance. For more information on the notation of tasks (i.e. A.1.1 to C.2.3), consult subsection 3.3.4 and Figure 3.3. **Figure B:** Represented as bar-graphs with standard error, the figure depicts time-based and spatial-based results over the eight evaluated sensory interface conditions C1 to C8. significance bars are also shown and denoted as: * $p < .05$, ** $p < .01$, *** $p < .001$ and finally n.s (or $p > .05$) standing for non significant. **Note for Figure B.1 to B.4:** For visualisation purposes, for all depicted measurement results entailing sensory combinations between monocular and stereoscopic vision, i.e. C1, C2, C3, C4 with C5, C6, C7, C8; are highly significant ($p < .001$). As a result, significance bars are visually omitted for those cases for enhanced visualisation purposes. **Figure B.1:** The time-based evaluation results, depicting the mean placement times (in opaque) and completion times (slightly transparent). **Figure B.2:** Spatial-based evaluation depicting the distance errors to the target. **Figure B.3:** Spatial-based evaluation depicting position accuracy results. **Figure B.4:** Spatial-based evaluation depicting rotation accuracy results.

Table 3.4: **Summary of the time-based evaluation metrics of the motor performance speed across different tasks with varying sensory input.** These entail placement and completion times across all eight interface conditions. Lower values correspond to better performance.

Objective Metrics (Time)				**Placement Time [s]**		**Completion Time [s]**	
	Vision	*Audio*	*Haptic*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*
C1	Monitor	Off	Off	14.27	±3.65	16.45	±3.91
C2	Monitor	Off	On	12.73	±2.98	14.70	±3.32
C3	Monitor	On	Off	13.14	±3.83	15.05	±3.42
C4	Monitor	On	On	13.00	±3.55	15.88	±3.15
C5	VRHMD	Off	Off	5.48	±2.10	9.22	±2.10
C6	VRHMD	Off	On	4.51	±1.57	7.58	±1.60
C7	VRHMD	On	Off	5.22	±1.77	8.92	±1.85
C8	VRHMD	On	On	4.47	±1.64	7.80	±1.71

quently, a Friedman's test was used, yielding a significant difference in mean placement as well as completion time across all eight multimodal interface conditions ($\chi^2(2) = 129.093$, $p < .001$) and ($\chi^2(2) = 131.093$, $p < .001$) respectively.

Further post-hoc analysis via the Wilcoxon Signed-Rank tests revealed a partial support of *Hypothesis H3*. More specifically, *Hypothesis H3.a* was supported as stereoscopic visual feedback stemming from the VRHMD (C5, C6, C7 and C8) accounted overall to significantly less placement as well as completion times than those conditions incorporating a monocular display monitor (C1, C2, C3 and C4) ($p < .001$ in both measurements). Furthermore, *Hypothesis H3.b* was partially supported as from the results somatosensory feedback contributed to additional significantly lesser placement and completion times only when paired with the VRHMD (C6 and C8) ($p < .05$ in both measurements). Finally, *Hypothesis H3.c* was not supported since auditory feedback did not contribute to a significant observable difference for either placement or completion times ($p > .05$ for both measurements). These results for both placement and completion times are detailed in Table 3.4 and depicted in Figure 3.5.B.1.

Spatial-Based Evaluation - Distance Error

For the spatial-based evaluation, distance error was initially recorded; defined as the distance from the object of interest to the indicated target location. The data was normally

distributed as revealed by a Shapiro-Wilk ($p > .05$). Consequently, a one-way RM-ANOVA with a GGC was used ($\epsilon = .381$). A highly statistical significance across the eight interface conditions was initially observed ($F(2.664, 63.950) = 90.463$,$p < .001$,$\eta_p^2 = .790$).

Post-hoc analysis revealed partial support of *Hypothesis H4*. More specifically, *Hypothesis H4.a* was supported as sensory interface conditions incorporating stereoscopic vision with the VRHMD (C5, C6, C7 and C8) accounted for significantly lower distance error ($p < .001$) than those compared to conditions with monocular vision from a display monitor (C1, C2, C3 and C4). In line with *Hypothesis H3.b* for the time-based evaluation, *Hypothesis H4.b* was partially supported since conditions incorporating somatosensory feedback showed further lower distance error to the target location, however, only within conditions incorporating stereoscopic visual feedback. More specifically conditions with the presence of haptic feedback (C6 and C8) revealed significantly lower target error to the target placement ($p < .05$), than conditions without (C5 and C7). In line with all the results presented up to this point, *Hypothesis H4.c* was again not supported since auditory stimulation did not contribute to any significant observable difference in lesser or higher distance error ($p > .05$). The distance error results are listed in detail in Table 3.5 and visually depicted in Figure 3.5.B.2.

Spatial-Based Evaluation - Position and Orientation Accuracy

Finally, position and orientation accuracy-based metrics were incorporated as part of the last spatial-based evaluation to fully understand how different sensory modalities affect spatial accuracy during motor tasks in full 3D space.

Consequently for spatial accuracy, position and orientation were recorded for each task across all eight multimodal interface conditions. Shapiro-Wilk Test in both instances yielded $p > .05$, thus signifying that the data was normally distributed. As such, a one-way RM-ANOVA was used; yielding in both position-accuracy and orientation-accuracy measurements a highly significant difference across the data, ($F(3.332, 79.962) = 174.488$,$p < .001$,$\eta_p^2 = .879$) with ($\epsilon = .476$) and ($F(3.139, 75.334) = 109.280$,$p < .001$,$\eta_p^2 = .820$) with ($\epsilon = .448$) respectively.

Contrary to *Hypothesis H1 to H4*, further post-hoc analysis revealed a full support of *Hypothesis H5*. More specifically *Hypothesis H5.a* was supported since interface conditions incorporating stereoscopic vision with the VRHMD (C5, C6, C7 and C8) accounted for significantly higher spatial accuracy both in position and orientation ($p < .001$), than those

Table 3.5: **Summary of the spatial-based evaluation metrics of the motor performance accuracy across different tasks with varying sensory input.** These entail distance error and translational as well as rotational accuracy across all eight interface conditions. For distance error, lower values correspond to better performance. For position and orientation accuracy higher values correspond to better performance.

Objective Metrics (Spatial)				**Dist. Error [cm]**		**Pos. Accuracy [%]**		**Rot. Accuracy [%]**	
	Vision	*Audio*	*Haptic*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*
C1	Monitor	Off	Off	19.12	±6.42	27.89%	±16.06	38.09%	±21.35
C2	Monitor	Off	On	17.58	±6.17	34.59%	±14.30	47.37%	±17.27
C3	Monitor	On	Off	16.15	±8.06	36.50%	±17.09	48.48%	±19.86
C4	Monitor	On	On	17.80	±8.61	35.22%	±16.11	45.44%	±16.97
C5	VRHMD	Off	Off	2.21	±1.08	77.75%	±5.93	87.55%	±7.30
C6	VRHMD	Off	On	1.65	±0.58	81.04%	±6.08	89.08%	±6.07
C7	VRHMD	On	Off	2.26	±1.37	77.65%	±7.75	87.27%	±6.65
C8	VRHMD	On	On	1.76	±0.70	80.97%	±6.16	90.01%	±5.67

conditions incorporating monocular vision from the display monitor (C1, C2, C3 and C4). In addition, *Hypothesis H5.b* was supported since conditions incorporating somatosensory feedback (C2 and C4 as well as C6 and C8) showed significantly higher spatial accuracy ($p < .05$), than those conditions with somatosensory feedback being absent (C1 and C3 as well as C5 and C7). Finally, full support of *Hypothesis H5.c* was inferred since sensory interface conditions incorporating only auditory stimulation (C3) resulted in greater spatial accuracy than those without (C1) ($p < .05$). From these findings, it can be inferred that spatial accuracy increases significantly when stereo vision is used and when paired with either sound or somatosensory or even both, rather than just relying on vision. This underlines the value of supplementing vision with somatosensory as well as auditory feedback showing a positive increase in spatial accuracy. The results of position and orientation accuracy are shown in Table 3.5 and visually depicted in Figure 3.5.B.3-B.4.

3.6 Discussion

From the experiments conducted, the inferred and analysed data, it can be inferred that the overall performance of participants increased by approximately 40% when using stereoscopic visual feedback from a VRHMD compared to monocular vision stemming from a

display monitor, with the percentage performance averaged over all subjective and objective metrics. An additional 10% increase in performance was observed when coupling somatosensory feedback, similarly as with stereoscopic vision, this was observed on all evaluation metrics incorporated in this study. Finally, auditory stimulation did not yield any significant observable difference via the implemented statistical testing conducted over any metric except a marginal increase of 5% towards spatial accuracy.

3.6.1 Implications of Results

These results suggest that visual feedback is the predominant sensory modality in the sensory-motor system of humans in the context of completing motor tasks of varying complexity. Such tasks are seen in virtual training and robotic teleoperation scenarios, in line with Rock and Victor (1964); Klatzky et al. (1998); Heilig (1992), supporting vision as a dominant sense. Moreover, the highly significant observable difference in performance as seen when comparing stereoscopic with monocular feedback via a VRHMD and a display monitor respectively, is primarily attributed to the superior depth information available via binocular vision as supported by current literature (Lampton et al., 1995; Witmer and Kline, 1998; Swan et al., 2015).

Most notably, it was observed that participants reported less perceived workload (measured with the NASA-TLX) and by extent lower cognitive demand when using stereoscopic vision via the VRHMD compared to the monocular view provided by the display monitor. While this contradicts the work of Brooks et al. (2017), this may be attributed to significantly higher amounts of induced "vection" when compared to the static scenario of upper-body motor movements investigated in this work. Nevertheless, further research is necessitated to confirm this finding and in particular, the different effects of lower and upper body movements may exhibit. Overall, these results are in line with existing work that stereoscopic vision is highly beneficial and leads to decreased mental demand when completing 3D motor tasks by increasing environmental awareness (Rosenberg, 1993; Scribner and Gombash, 1998), with other studies contradicting these findings and reporting higher workload demand than others (Johnson et al., 2015). More investigation on the effects of VRHMDs would aid in solidifying further these findings.

From the results it can also be inferred that somatosensory feedback leads to increased performance which is supported by some studies (Brickler et al., 2018), however, contradicts others (Frid et al., 2018). The latter study found no significant effect of somatosensory

feedback in a virtual throwing task. However, the large number of design options available when simulating somatosensory feedback as well as the multitude of different haptic devices, can likely result in different observations and conclusions when studying different types of stimuli.

Furthermore, additional difficulties are presented when attempting to infer the efficiency of haptic devices due to the "Uncanny Valley of Haptics" phenomenon proposed by Berger et al. (2018). As with the original term of the "Uncanny Valley" (Mori et al., 2012), when simulating increasing resolutions of somatosensory feedback without the corresponding level of resolution and stimulation from other senses of bi-modal feedback or more, will not contribute to a guaranteed increase in performance. This implies that the resolution of all included stimuli from other simulated modalities within multimodal interfaces, which by definition would at least include two sensory states, needs to be similar. To which extent this can be quantified remains an ongoing challenge (Berger et al., 2018). Nevertheless, it can be supported by these results, in line with Berger et al. (2018), the suggested phenomenon that somatosensory feedback increases performance primarily during increasingly corresponding visual stimulation. More specifically, from the results it can be observed that increased perceived *(subjective)* and measured *(objective)* performance was primarily observed when coupling somatosensory feedback with increased visual stimulation stemming from a VRHMD providing stereoscopic rather than monocular vision from a display monitor. These results suggest that the "Uncanny Valley of Haptics" may indeed be a phenomenon that should be accounted for when designing multimodal interfaces and ensuring to certain extents that the included modalities attain similar levels of resolution.

Minimal evidence was revealed regarding the benefits of auditory feedback. It was inferred that while auditory feedback did not contribute towards task performance in most of the incorporated metrics, spatial accuracy did increase in the bi-modal audiovisual condition compared to solely relying on vision ($p < .05$). Furthermore, it was observed that workload demand, via the NASA-TLX, marginally decreased when auditory feedback was presented rather than being absent, however, not to a significant degree ($p = .056$). In contrast to the findings of this work, a significant difference was found in the work of Nagai et al. (1999). It is hypothesised that this could be due to a bi-product of the increase in performance when switching from mono to stereo vision, potentially overshadowing the contribution of audio. As seen from the results of this work and in particular the perceived

subjective performance reported from participants, auditory feedback was not deemed to be necessary in contrast to stereo vision being the most apparent in increased perception. Nevertheless, spatial accuracy appears to increase marginally, yet not to a significant difference when coupling auditory feedback, potentially implying that tri-modal feedback i.e. vision, audition and haptics are beneficial instead of presenting any inverse or counter-productive performance.

3.6.2 Summary of Main Findings

The main findings of this research and the subsequent derived design implications include:

- The addition of more sensory modalities within multimodal interfaces is correlated to improved task performance;
- Vision dominates within audiovisuohaptic interfaces and all possible combinations therein, rendering a notably higher sensory contribution and by extension resulting in better performance when switching from monocular to stereoscopic vision;
- Prioritisation of visual (stereoscopic over monoscopic), somatosensory (presence over absence) and then auditory stimulation (presence over absence) should be considered during manipulation tasks;
- The effectiveness of multimodal interfaces is scenario-specific, this research explored it in the context of manipulation and did not account for lower or upper body navigation;
- Increasing motor task complexity lowers effectiveness as expected, however, is not proportional for all sensory combinations.

The results of this chapter provide new evidence for the untested hypothesis of Sigrist et al. (2013). In particular, novel evidence was provided that audiovisuohaptic interfaces incorporating a stereoscopic VRHMD over a monocular monitor contribute to the highest observable motor performance. Visual stimulation with either the VRHMD for stereoscopic vision or a generic monocular monitor, are the most commonly used visual devices in related work concerned with robotic teleoperation and simulation training (Lipton et al., 2018; Triantafyllidis et al., 2020; Triantafyllidis and Li, 2021b). Higher motor performance is additionally observed with visuohaptic and less closely by audiovisual interfaces. Figure 3.6 visually represents a cone-like illustration of the effectiveness of

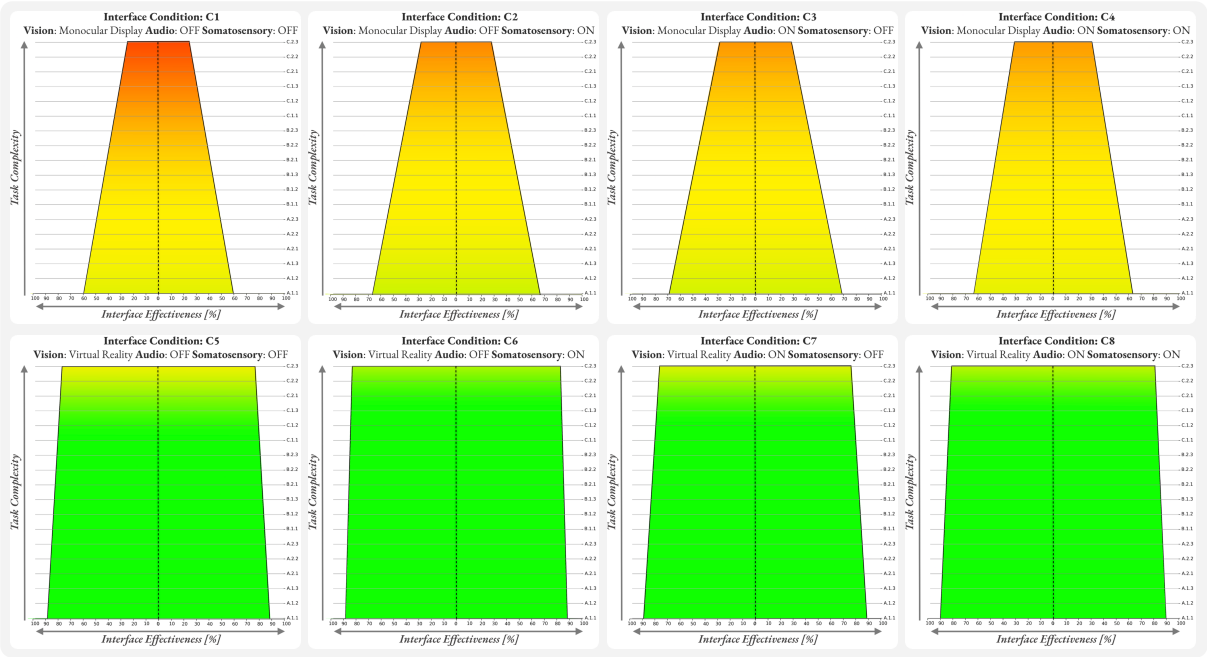

Figure 3.6: **Overall sensory interface effectiveness through linear regression across all measurements.** The figure depicts the collective effectiveness of each sensory interface condition (C1 to C8) across all sensory measurements. Eight distinctive cones are depicted, corresponding to the eight interface conditions. The cone-like illustrations of these are inspired by Sigrist et al. (2013), providing a similarly looking plot to aid inter-study comparability of the new findings of this work. **X-Axis:** The horizontal x-axis represents the interface effectiveness and in particular the width of the shape. The wider the shape the higher the effectiveness. Colouring is added to the shape to aid visualisation and is proportional to the cone's width, with interface effectiveness ascending from red to green. **Y-Axis:** The vertical y-axis represents the motor task complexity, with higher spatial complexity ascending from bottom to top. Notations of the specific tasks A.1.1 to C.2.3 are made for each cone representation. **Notes:** The overall effectiveness is calculated linearly, specifically, measurements are weighted $(1-1/V_{max})$, whereby V_{max} is the maximum limit of the measurement. The data points from the scatter plot have been line-fitted through linear regression to visualize a cone-like illustration and are hence mirrored. The width of the cones represents the effectiveness of the sensory interface while the height of the cone the effectiveness of that interface at the specific spatial complexity of the task studied. For more information on the notation of tasks (i.e. A.1.1 to C.2.3), consult Section 3.3.4 and Figure 3.3

Table 3.6: **The summary of the hypotheses and their support entailing a plethora of different sensory stimuli and their effects on motor performance across varying spatial complexity tasks.** Character abbreviations are as follows – Y: Yes. P: Partial. N: No.

Hypothesis	**Support**	**Detailed Description**
H1: Lower perceived workload	Partial	(a) Y (b) P (c) N
H2: Higher system usability	Partial	(a) Y (b) P (c) N
H3: Less task time	Partial	(a) Y (b) P (c) N
H4: Less distance error	Partial	(a) Y (b) P (c) N
H5: Higher placement precision	Full	(a) Y (b) Y (c) Y

(a) Vision via the VRHMD compared to monitor

(b) Haptic feedback compared to absence

(c) Sound feedback compared to absence

*P: Partial; only effective when paired with stereo VRHMD

the different possible combinations of sensory feedback stemming from audiovisuohaptic multimodal interfaces, across increasing levels of motor task complexity. These cone-like illustrations closely resemble, for increased inter-study comparability, the visual representation from the work of Sigrist et al. (2013). Figure 3.6 provide, to the best of the author's knowledge, novel evidence of such untested, to this date, multimodal combinations. From Figure 3.6 it can be consequently inferred that tri-modal feedback with the addition of audio and somatosensory feedback (C8) contributes the highest towards task performance, even amongst the most complex motor tasks. It can further be concluded that the addition of more sensory modalities is correlated to improved performance in line with existing work (Burke et al., 2006). Moreover, when combining in order of significance: visual (stereoscopic over monoscopic), somatosensory (presence over absence) and auditory (presence over absence) feedback, the contribution towards overall task performance in 3D 6DoF manipulation is maximised.

3.6.3 Limitations and Future Work

In this chapter, part of the initial aim of this PhD thesis was presented. In particular, how different multimodal sensory modalities and all their possible different sensory combinations affect the human sensory-motor system, reflected on motor performance in 3D motor tasks entailing the manipulation of objects of varying spatial complexity.

The limitations and future work of this work include:

- ***Overloaded Sensory Stimuli*** – In the experiments conducted, non-overloaded visual stimuli were compared. While it was comprehensively studied how stereoscopic vision compares to monocular vision, it would be of potential importance to also study how other sensory modalities may prove to be a promising supplementary sense when visual information is overloaded. Studying for example how auditory and somatosensory feedback can compensate for overloaded visual information, could in essence provide further insight on their importance and contribution towards human performance.
- ***Monocular and Stereo Rendering on a VRHMD*** – Moreover, two common visual feedback devices in current work were compared, namely a VRHMD and a generic display monitor. For future work, it would be important to solely limit this to one single device and instead compare stereoscopic and monocular states. For example, rendering stereoscopic or monocular vision at a given time solely on one device (such as a VRHMD) may shed additional light on how those modalities compare against each other. While this would allow visual feedback to be explored from a theoretical approach, the investigation conducted in this study explored this from a practical approach, since VRHMDs and monocular display monitors are the predominant visual devices used in existing work on robotic teleoperation and interactions within virtual environments (Lipton et al., 2018; Triantafyllidis and Li, 2021b; Triantafyllidis et al., 2020).
- ***Sensory Design Decisions of Multimodal Interfaces*** – Additionally, the sensory design decisions when evaluating multimodal interfaces are of importance (Sigrist et al., 2013). For instance, comparing different types of somatosensory feedback devices such as vibration or force feedback gloves and whether those are used during collisions with virtual objects or as the means of indicating imminent collision are all likely going to lead to different results and conclusions. Consequently, the design decisions as well as the compared hardware of each sensory channel will affect task efficiency and the implications of those results. In this work, the investigation of such modalities was limited to purely the presence or absence of both auditory and somatosensory feedback, instead of comparing terminal or concurrent feedback for example.
- ***Latency Induced Sensory Feedback and Synchronisation of Modalities***

– Lastly, another potentially interesting research path is how latency may affect different sensory modalities and how certain ones may prove to be promising supplementary alternatives. The work in this chapter limited its investigation to zero or minimal latency during the conducted experiments. Current work supports that time delays are correlated to simulator sickness, which is a real-world problem in teleoperation and further aggravated in wireless technologies (McGlynn and Rogers, 2017), especially entailing MR devices (Triantafyllidis et al., 2020). Future work could explore how latency affects human motor performance and more broadly the sensory-motor system. This could furthermore be extended by studying how synchronisation within different combinations of multimodal interfaces can perhaps lead to decreased or even increased performance with the hopes of quantifying this relationship and generalising a set of guidelines when designing such interfaces.

Additional details of the future work with concluding remarks are presented in Chapter 7. Within this study and to the best of the author's knowledge, a thorough comparison of audiovisuohaptic multimodal interfaces and all their possible combinations was conducted. Invaluable insights were gained on which modalities contribute to increased motor task performance during complex in 3D with 6DoF motor tasks, as commonly seen in robotic manipulation during teleoperation, MR-based settings and more broadly in interactions within VEs.

Having now gained a better understanding of the human-sensory motor system and how it can be harnessed in the context of upper-body motor tasks with different sensory stimulation and interfaces, Chapter 4 will explore the intriguing idea of quantifying these biomechanical movements and actions.

Chapter 4

Quantifying and Measuring Human Motor Actions and Movements

In the previous chapter of the thesis (Chapter 3) the first goal of the PhD project was presented. An investigation was conducted on how different sensory feedback states and their combinations affect the sensory-human system of humans, reflected on motor performance in 3D space with 6DoF across varying spatial difficulty tasks.

Nevertheless, measuring human performance is a complex phenomenon, especially when attempting to quantify it Triantafyllidis and Li (2021b). In Chapter 3 a plethora of different time and spatial-based evaluation metrics were incorporated, similarly to Triantafyllidis et al. (2020). The primary limitation when attempting to evaluate via the use of such a multitude of different metrics is decreased inter-study comparability amongst related work (Triantafyllidis and Li, 2021a). This is due to the absence of a standardised evaluation tool. This chapter addresses this gap and derives a novel metric capable of quantifying such biomechanical movements from the human sensory-motor system.

Parts of this chapter are published as: E. Triantafyllidis, W. Hu, C. McGreavy and Z. Li, "Metrics for 3D Object Pointing and Manipulation in Virtual Reality: The Introduction and Validation of a Novel Approach in Measuring Human Performance". In: *IEEE Robotics & Automation Magazine*, Volume: 29, No. 1, Pages: 76-91, March 2022, DOI: 10.1109/MRA.2021.3090070.

Parts of this chapter are published as: E. Triantafyllidis and Z. Li, "The Challenges in Modeling Human Performance in 3D Space with Fitts' Law." In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. May 2021. Association for Computing Machinery, New York, NY, USA, Article 56, 1–9. https://doi.org/10.1145/3411763.3443442

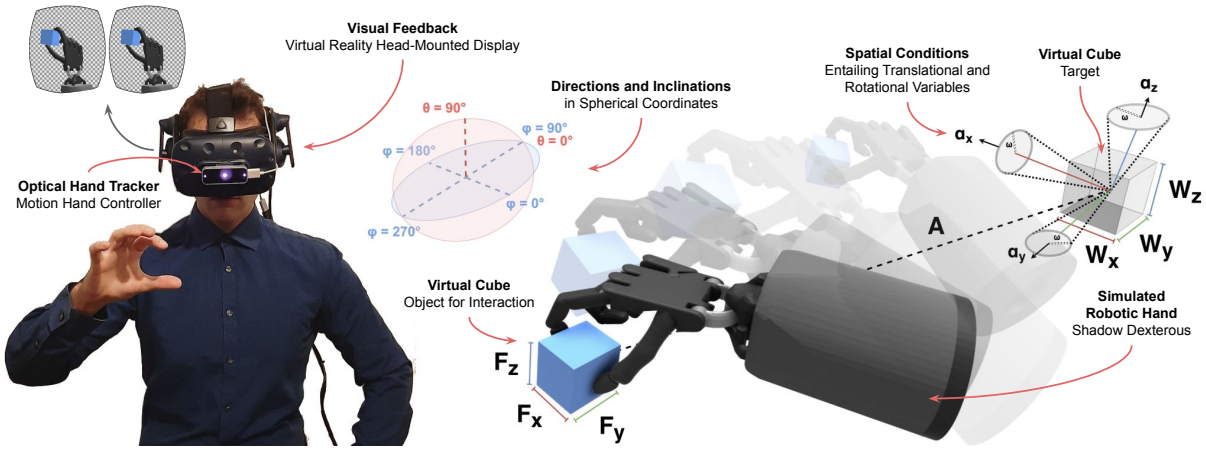

Figure 4.1: **An operator performing motor tasks of varying complexity in full 3D space.** A representative example of the work of this chapter. The figure depicts a human operator performing motor tasks of varying complexity to satisfy a plethora of different spatial criteria. The simulated task depicted entails various translational and rotational motions with the presence of directional and inclination variations. Such motions are commonly seen in robotics and virtual environments, particularly those incorporating MR technologies. It is important to classify which of these spatial variables have a notable effect on the human sensory-motor system and in particular on motor performance. Ultimately, this will allow for the ability to measure and quantify such bio-mechanical movements via the derivation of an appropriate metric. This chapter addresses this gap.

In this chapter and as part of the second aim of this thesis, an extensive comparison of existing state-of-the-art human motor evaluation metrics will be conducted, with the aim of deriving a more general and standardised formulation that can encapsulate multiple measurements under a single model. Paul Fitts' original human predictive model – short for **Fitts' Law**, represents the gold standard as a human predictive model (Triantafyllidis and Li, 2021a; MacKenzie, 1992). Most notably due to its significant popularity and its ability to encapsulate multiple measurements under one formulation, the law represents a promising basis for deriving an extended formulation (Triantafyllidis and Li, 2021a). To this end, an emphasis is directed to Fitts' law. As part of the experimental evaluation, a set of four, with increasing spatial complexity motor tasks in 3D space will be utilised to comprehensively evaluate the effectiveness of existing human predictive models based on Fitts' original law including popular extensions of it. As part of this evaluation process, a thorough understanding can be gained of the entailed spatial variables as seen in higher-dimensional motor movements with the use of 6DoF (Triantafyllidis et al., 2020). Ultimately, as part of this chapter, this work will be deriving a novel evaluation metric based on Fitts' law and its extensions, that can collectively capture the complex spatial

variables and motor movements associated with full 3D space. Implications for solidifying the metric and future work are presented in the final chapter (Chapter 7). Figure 4.1 visualises the simulation setup and the studied motor tasks that were subsequently analysed for the derivation of a novel metric as part of the second contribution of this thesis.

Chapter Outline – This chapter is organised in several sections, Section 4.1 first presents a short introduction and builds upon the motivation and the necessity of an evaluation metric covering fully the 3D domain. Section 4.2 presents the methodology, system apparatus and experimental design accommodating this work. Next, in Section 4.3 the experiments are presented as well as the statistical procedures that were used to evaluate the subsequent acquired data. In Section 4.4 the results are summarised in particular how existing evaluation metrics explain the spatial variations investigated in the experiments and which variables contributed the highest towards task difficulty. Section 4.5 outlines the derivation of a novel metric extension as the result of this work, covering the full 3D domain, with each spatial variable of the metric detailed and statistically analysed. Finally, in Section 4.6 the results are discussed and reflected on the state-of-the-art while also comparing the derived metric against existing ones. Finally, the limitations and future work are discussed briefly, with more expansion in Chapter 7.

4.1 Introduction and Background

As inferred from the literature review in earlier chapters of this thesis, assessing human performance is of high importance when deriving new formulations that surround users (in VEs) and operators (in teleoperation). Ultimately, to comprehensively understand the sensory-motor system of humans and the underlying biological mechanisms that surround this intricate system and render humans proficient in meeting complex environmental demands, the biomechanical movements generated by this system need to be measured and quantified (Triantafyllidis and Li, 2021a; Gonzalez-Franco and Peck, 2018; Ranscombe et al., 2019; Kilteni et al., 2012). The necessity of such a metric has exponentially increased in recent years, primarily attributed to the notable growth of realistic with high-fidelity simulators, networking and MR technologies (Triantafyllidis et al., 2020; Drewes, 2010). For more details regarding a thorough literature review, consult Section 2.2.

The Necessity and Importance a Human Performance Metric – Evaluating and measuring user performance in VEs has always been a longstanding research goal (Savage-Knepshield et al., 2016; Triantafyllidis and Li, 2021a). However, even to date and to the best of the author's knowledge, a compelling standardised metric for 3D object manipulation and general 3D interactions does not yet exist. The absence of a standardised metric severely limits inter-study comparability between the results of different studies. This is primarily due to the multitude of different metrics researchers can use, which aggravates generalisation and ultimately conclusions according to (Steinfeld et al., 2006). Hence, progress towards the endeavour of deriving a standardised formulation is still "scattered" amongst related studies, which limits solidifying a concrete and established metric (Blackman et al., 1992; Gawron et al., 1990; Drewes, 2010). Current extensions often disregard important spatial aspects that one would expect in full 3D space, rendering the establishment of more generalised patterns difficult (Triantafyllidis and Li, 2021a).

Fitts' law: A Promising Basis for a Metric– To propose a higher dimensional metric for assessing human performance, in this work, Paul Fitts' original predictive model is investigated, short for "Fitts' law" (Fitts, 1954; Fitts and Peterson, 1964). Proposed in 1954, the law has been extensively used in HCI and ergonomics research and soon grew to become the standardised approach of measuring human movement time based on simple spatial settings, which still represents the gold standard as a human predictive model (Triantafyllidis and Li, 2021a; MacKenzie, 1992). This popularity of the law is attributed to the advantage of measuring human performance in a time-based approach based on spatial data, effectively combining both time and spatial-based metrics under one single formulation and aiding inter-study comparability, as supported by Triantafyllidis and Li (2021a).

Extensions Derived from the Law – Fitts' formulation was originally used for 1D translational movements (Fitts, 1954), subsequently extended to 2D tasks (MacKenzie, 1992; Hoffmann, 1995) with its applicability further highlighted in 2D rotational tasks (Kondraske, 1994; Kulik et al., 2020; Stoelen and Akin, 2010). In more recent work, Fitts' law was also extended to some extent in 3D space, however, limited to translational tasks, with numerous reformulations and extensions (Barrera Machuca and Stuerzlinger, 2019; Cha and Myung, 2013; Murata and Iwase, 2001). Unfortunately, most studies

to date attempting to extend the law have either disregarded important spatial aspects associated with full 3D space entailing higher DoFs or limited their findings to very specific settings and study manifestations.

The Absence of an Appropriate Full 3D Extension of the Law – Previously proposed 3D metrics extending Fitts' law to higher dimensional space, have disregarded combining translational and rotational movements (Barrera Machuca and Stuerzlinger, 2019; Cha and Myung, 2013; Murata and Iwase, 2001), assessing directional (Kulik et al., 2020; Stoelen and Akin, 2010) or inclinational variations (Barrera Machuca and Stuerzlinger, 2019; Kulik et al., 2020; Stoelen and Akin, 2010; Murata and Iwase, 2001) in one exhaustive study. However, all of the above spatial settings are of utmost importance when considering the entirety of 3D space with 6DoF which is commonly inherent in human movement according to Triantafyllidis and Li (2021a). Moreover, most studies to date only focused on pointing tasks as per the original formulation of Fitt's law. However, it would be invaluable to test and investigate whether the law can also be applied to manipulation tasks, i.e. with the incorporation of grasping and physical properties one would expect in real life (Triantafyllidis and Li, 2021a). Investigating whether an extension of the law can be derived covering both the pointing and the manipulation domain will notably aid subsequent studies in the domain of 3D user interfaces and robotic teleoperation as well as MR simulators (Triantafyllidis et al., 2020). Consequently, as it can be inferred and as supported by the works of Drewes (2010); Triantafyllidis and Li (2021a), collectively combining these spatial variables is of necessity for the derivation of an appropriate full 3D extension of Fitts' law with the hopes of standardising to a great extent a human performance metric. Figure 4.1 illustrates the common full 3D motor tasks entailing 6DoFs commonly encountered in human motion. Table 4.1 reports in detail the most widely used model extensions based on Fitts' law, their characteristics and ultimately their applicability in full 3D space. For more details regarding the shortcomings of Fitts' original law for higher-dimensional space, consult Section 2.2.2.

Research Aim and Objective – From the above it can be inferred that due to the law's inherent characteristic of (i) fusing two different types of metrics under a single model and (ii) its popularity and subsequent research focus, it would be invaluable to investigate the law's applicability and potential to be extended in higher dimensions (Triantafyllidis and Li, 2021b). By proposing a 3D extension of Fitt's law, which still is to date a gap

in existing work; this chapter hopes to address that problem by increasing comparability and generalisation towards the 3D manipulation domain (Triantafyllidis et al., 2021).

4.2 Methodology

A large part of the simulation environment and the apparatus used for this work was based on Chapter 3, outlined in more detail in Section 3.3. Figure 4.1 visually depicts the overall work of this chapter, entailing the system setup, technologies used and the motor tasks studied in order to derive a novel metric.

4.2.1 System Apparatus

The overall system apparatus was to a great extent similar to that presented in Chapter 3 (shown also in Figure 3.2). In brief, the simulation engine Unity3D was used as the primary simulator accommodating all necessary hardware and software. To allow for the assessment of motor movement within the virtual simulation, the LMHC and the accompanying SDK were integrated to allow for optical hand tracking. The LMHC's two infrared cameras with a sample rate of 120Hz and a FoV of 135° were deemed sufficient as the means of the input interface. Furthermore, to ensure optimal hand-tracking performance, the lighting conditions in the physical location of participants were consistent. Moreover, the operational space was limited to about 100cm as the upper maximum reaching bounds from the chest of users, in accordance with the limitations of the optical tracking capabilities of the LMHC. Lastly, a low-pass filter with a cutoff frequency of 10Hz was implemented and incorporated into the LMHC to reduce optical inconsistencies and ultimately noise during retargeting, ensuring continuity and robustness.

In terms of physics, the integrated physics engine PhysX in Unity3D was used and the physics simulation time-step was set at 1,000Hz to ensure robust and stable performance of the associated physical properties one would expect from the real world such as forces and friction coefficients.

In terms of visual feedback, high-resolution displays were used to limit distance overestimation and degraded longitudinal control which are both known issues in VEs in accordance to the related works of Barrera Machuca and Stuerzlinger (2019); Triantafyllidis and Li (2021a). Hence, similarly to Chapter 3, the VRHMD HTC Vive Pro was chosen for this purpose, entailing a display resolution of 2,880 x 1,600 pixels and a 110° FoV at

Table 4.1: **Summary of the most widely used extensions and models of Fitts' law.** In these tables, the model equations and the model characteristics in Table 4.1a and Table 4.1b are reported respectively.

Model Extensions and Equations		
Human Performance Models	**Model Formulation**	
	MT	ID_t / ID_r*
Fitts (1954)	$MT = a + b \cdot ID$	$ID_t = log_2\left(\frac{2A}{W}\right)$
MacKenzie (1992)	$MT = a + b \cdot ID$	$ID_t = log_2\left(\frac{A}{W} + 1\right)$
Hoffmann (1995)	$MT = a + b \cdot ID$	$ID_t = log_2\left(\frac{2A}{W+F}\right)$
Welford (1968)	$MT = a + b \cdot ID$	$ID_t = log_2\left(\frac{A}{W} + 0.5\right)$
Murata and Iwase (2001)	$MT = a + b \cdot ID$	$ID_t = log_2\left(\frac{A}{W} + 1\right) + c \cdot \sin\theta$
Cha and Myung (2013)	$MT = a + b \cdot \theta_1 + c \cdot \sin\theta_2 + d \cdot ID$	$ID_t = log_2\left(\frac{2A}{W+F}\right)$
Stoelen and Akin (2010)	$MT = a + b \cdot [ID_t + ID_r]$	$ID_{t/r} = log_2\left(\frac{A_t/\alpha_r}{W_t/\omega_r} + 1\right)$
Barrera Machuca and Stuerzlinger (2019)	$MT = a + b \cdot ID + c \cdot CTD$	$ID_t = log_2\left(\frac{A}{W} + 1\right)$

(a) The table illustrates the original equations as defined by the respective authors. $ID_{t/r}$: Index of difficulty of translation (t) or rotation (r).

Model Extensions and Attributes					
Human Performance Models	**Model Characteristics**				
	Based On	Space	Dir.*	Inc.**	Depth
Fitts (1954)	N/A	2D	No	No	No
MacKenzie (1992)	Fitts	2D	No	No	No
Hoffmann (1995)	Fitts	2D	No	No	No
Welford (1968)	Fitts	2D	No	No	No
Murata and Iwase (2001)	MacKenzie	3D	Yes	No	No
Cha and Myung (2013)	Murata and Iwase; Hoffmann	3D	Yes	Yes	No
Stoelen and Akin (2010)	MacKenzie	3D	No	No	No
Barrera Machuca and Stuerzlinger (2019)	MacKenzie	3D	No***	No	Yes

(b) The table details the characteristics of the model extensions and whether these cover important spatial characteristics to model full 3D performance. **Dir.*:** Directions. **Inc.**:** Inclines. **No***:** Effects were investigated but no formulation or model extension was performed. **N/A:** Not applicable.

90Hz. Head rotations within the VE were made possible with the included and built-in photosensors of the VRHMD. For all experiments hereinafter presented, the LMHC was fixed on the front of the VRHMD.

For hand control, the virtual Shadow Dexterous robotic hand was implemented in the simulation via the Robot Operating System (ROS), in particular the ROS# plugin, allowing the import of physics models of robots and objects (Unified Robot Description Format (URDF)). Figure 4.2.A, similarly to Chapter 3 and in particular Figure 3.2.B.1.

4.2.2 Hand Motor Control and Input Interface

For the experiments, it was crucial to design a stable and realistic physic-capable hand motor control interface. In Figure 4.2.B the user's hand movements are visually depicted and more specifically how these were mapped onto an anthropomorphic hand within the simulation, the Shadow Robotic hand. A total of 6DoF were allowed within the simulation to match the motions of real-world robotic teleoperation and VR-based interactions and hence users could move freely around the VE in all axes entailing both translation and rotation. The Cartesian hand movements and joint positions from the participant's hand were recorded via the LMHC and thereafter mapped onto the simulated robotic hand within the VE allowing for efficient and realistic tele-manipulation. As described in detail in Section 3.3, a similar hand-retargeting method as in the published work of Triantafyllidis et al. (2020) was used. In particular, the joint positions of the user's fingers were obtained by calculating the angle θ between a joint $\overrightarrow{b_{i-1}}$ and its parent joint $\overrightarrow{b_i}$ optically tracked by the LMHC. Thereafter, the resultant angle θ from the user's finger was incorporated in a joint PD controller to achieve the desired re-targeting joint motions from the user's hand to the simulated robotic one ensuring a smooth and stable transition. The re-targeting approach is formulated mathematically as:

$$\begin{aligned} \theta &= \arccos\left(\frac{\overrightarrow{b_i}\cdot\overrightarrow{b_{i-1}}}{\left\|\vec{b_i}\right\|\left\|\overrightarrow{b_{i-1}}\right\|}\right), \\ \tau_i &= K_{P_i}\cdot(\theta_{d_i}-\theta_{c_i}) - K_{D_i}\cdot\dot{\theta}_i, \end{aligned} \tag{4.1}$$

whereby θ represents the desired angle for the virtual robotic hand to match. Moreover, τ represents the torques applied to each joint i. θ_{d_i} and θ_{c_i} are the desired angles from the human hand from the LMHC and the current angle of the virtual hand joint in the simulation, respectively. Finally, the variable $\dot{\theta}_i$ is the measured velocity of the virtual hand

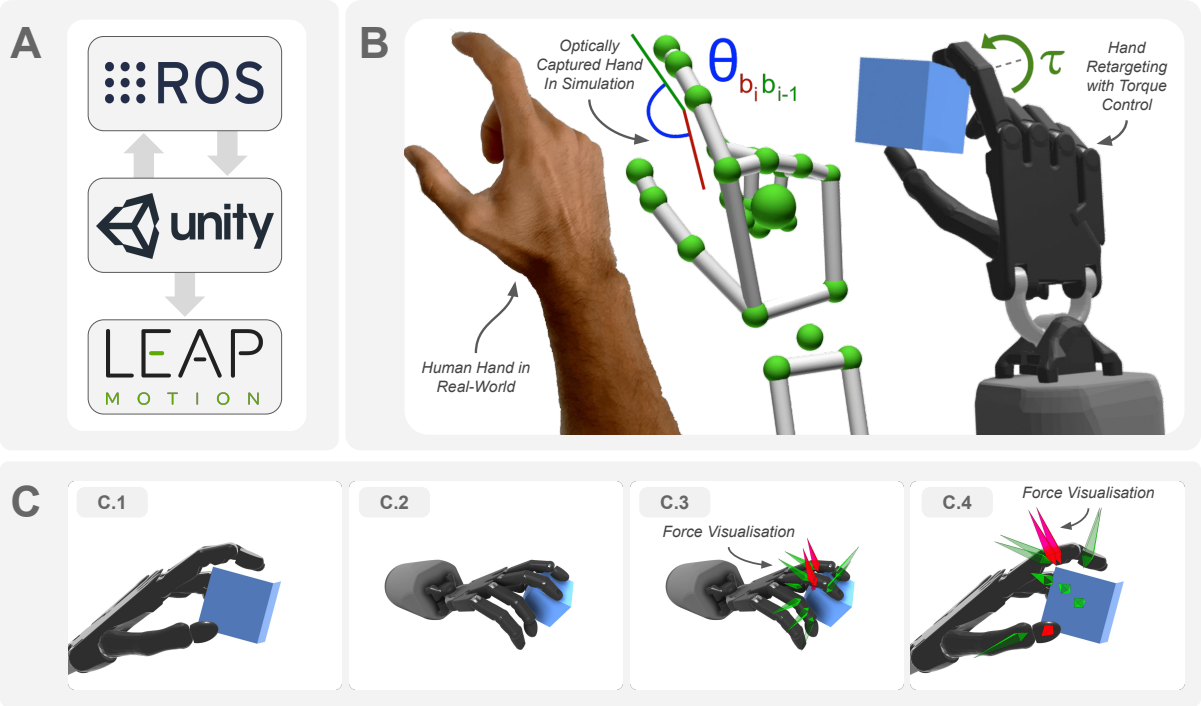

Figure 4.2: **Simulation, hand control and physics overview.** The figure depicts the simulation overview, the employed hand control and re-targeting approach as well as the physics visualisation during grasping. **Figure A:** The simulation overview and all entailed software. The simulation setup is retained to a great extent from Chapter 3 and in particular Figure 3.2.B.1. **Figure B:** The employed hand control, similar to Figure 3.2.A.2 from Chapter 3, with the difference of applying torque instead of velocity control to the robotic hand. **Figure C:** Depicts the grasping of objects for the manipulation task cases and in particular the involved physics. **C.1 and C.2:** Grasping of the cube target. **C.3 and C4:** Grasping of the cube target with force visualisation, with increased force from green to red respectively. **Note:** The sample physics engine and settings entailing all physical grasping of objects in the simulation were used for both Chapter 3 and Chapter 4 for consistency.

for computing the damping torque. Figure 4.2.B visually depicts the hand-retargeting approach and Figure 4.2.C the forces exerted in the physics simulation.

Finally, a velocity control signal was applied to the robotic hand based on the real hand position of the participant and orientation to control the six in total DoFs of the robotic hand in the simulated environment. The collisions between the robotic hand and the objects in the simulation were realised via the NVIDIA PhysX 4.0 engine within Unity3D. The original joint limits and colliders of the virtual hand were retained, as specified in the URDF file of the Shadow robot hand to ensure optimal and realistic actuation as one would expect from the real hardware.

4.2.3 Task Design and Motor Assessment

The experiments conducted in this work entailed different types of motor task movements with progressively higher spatial difficulty. More specifically, the experiments entailed four in total motor tasks. The motor tasks and all investigated spatial variables entailing translational and rotational movements are outlined in full detail in Table 4.2.

In each of those motor tasks, the participants were asked to move an object from a constant pre-defined starting position to a variable target location. The object manipulated by users was presented as a solid blue cube and the target location as a transparent cubic volume. The object shape used in this work entailed a 3D cube to allow the assessment of rotational variations of motor tasks, which is contrary to the use of a sphere or the index fingers of participants as seen commonly in related work and limiting their investigation in purely translational settings (Barrera Machuca and Stuerzlinger, 2019; Cha and Myung, 2013; Murata and Iwase, 2001). Moreover, due to the 3D cube's identifiable orientation and as one of the most basic 3D shapes it presented a suitable choice to assess both translational and rotational motor tasks. In regards to rotational motions as part of the motor assessment, participants were instructed to match the sides of the cube with that of the cube target to achieve the most parallel placement possible. While the main limitation of using a cube was that only four, in essence, "correct" rotations could be assessed and the range of rotational degrees entailed a maximum of 45 degrees, the cube nonetheless represented the dominant and most widely used 3D shape in current work (Triantafyllidis et al., 2020). The assessment of rotational movements in related work has indeed been conducted and investigated. However, studies to date investigated rotational variations, limited to 2D movements without extensions to the 3D space (Kulik et al., 2020; Stoelen and Akin, 2010). All targets in the subsequent presented motor task experiments were arranged in the spherical coordinate system with the object at the centre. Figure 4.3.A and Figure 4.3.B illustrate movements in 2D and full 3D space respectively.

In terms of task success, an intersection of the cube object with its transparent cube volume target counterpart was sufficient for task progression so long the translational and/or rotational requirements were met. These requirements varied depending on the task type and difficulty. As the means of visual indication of task succession, the transparent target would turn green, indicating success and progression, as shown in Figure 4.4. In more detail, for translational types of motor movement tasks, a 50% overlap with the target was considered a success, in line with Fitts' original experiment. For rotational motor tasks,

Table 4.2: **The investigated spatial variables entailed in higher-dimensional space with 6DoFs to comprehensively evaluate which have the most profound effect on motor performance.** The table details the parameter settings of all spatial types as well as variations investigated in this work. **Note on Metrics:** Variables F, W and A are in centimeters (cm). Variables ϕ, θ, α and ω are in degrees (°).

Variables		**Variations Investigated**			
		Motor Task 1	Motor Task 2	Motor Task 3	Motor Task 4
F	Object Size	(3, 4, 5)	(5)	(4, 5)	(4)
W	Target Width	(5, 7.5, 10, 12.5)	(5, 10)	(5, 10)	(4, 8)
A	Target Separation	(12, 24, 36, 48)	(12, 24)	(0)	(12, 24)
ϕ	Direction Angle	(90)	(0, 90, 180, 270)	(0)	(0, 90)
θ	Inclination Angle	(0)	(15, 30, 45)	(0)	(15, 30)
α	Angular Distance	(0)	(0)	(15, 30, 45)	(30, 45)
ω	Angular Tolerance	(0)	(0)	(2.5, 5, 7.5, 10)	(7.5,15)
Variations x (Repetitions)		48x(5)=240	48x(5)=240	48x(5)=240	64x(4)=256

the target rotation, α, needed to be matched within a certain rotation tolerance ω in all axes in 3D space. Figure 4.4 and Table 4.3 visually depict and detail the mathematical equations that needed to be satisfied for a task to be classified either as a success or an error respectively.

Pointing and Manipulation – In addition to varying all possible spatial variables in 3D space entailing translational and rotational variations, pointing and manipulation types of tasks were investigated, as these represented the most common types of interactions in both VEs (Barrera Machuca and Stuerzlinger, 2019) and teleoperation (Triantafyllidis et al., 2020; O'Hara, 1987). Foremost, pointing types of interactions were investigated as these closely follow Fitts' original experiment. As identified from the literature review, humans actively point towards targets in their daily lives; representing pointing as the dominant type of interaction, particularly in 3D user interfaces and their resemblance with peg insertion tasks as seen in teleoperation (O'Hara, 1987). Secondly, to cover a current limitation in literature when utilising and investigating Fitts' law for robotics, manipulation tasks were studied. These types of task interactions, unlike pointing, entailed physically grasping the items with realistic physics as seen in teleoperation (Lipton et al., 2018; Triantafyllidis et al., 2020; Triantafyllidis and Li, 2021a).

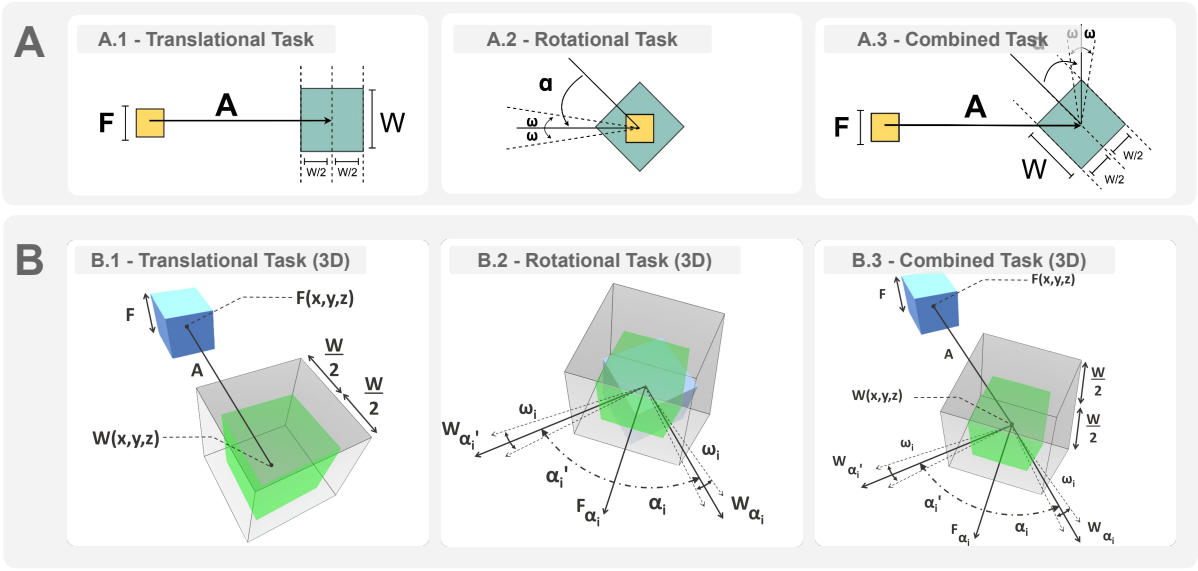

Figure 4.3: **Overview of low-dimensional and high-dimensional representations of different motor tasks in 2D and 3D space.** The figure depicts 2D and 3D tasks. **Figure A:** Represents the simplest 2D setting of a pointing task that can be extended towards manipulation as well. **A.1:** The simplest setting of a translation task, entailing a single-line movement to the target. **A.2:** A rotational task necessitating the rotation of the object to that of the target, entailing a certain rotational tolerance. **A.3:** Combined translational and rotational task, necessitating the spatial satisfaction of the variables entailed in A.1 and A.2. **Figure B:** The equivalent to Figure A, full 3D task in 6DoF. Satisfying the entailed spatial requirements of this task is notably higher than the 2D equivalent due to increased dimensions and by extent DoFs. **B.1:** Translational task necessitating a 3D translational movement. **B.2:** Rotational task necessitating a 3D rotational movement, entailing a certain rotational tolerance along all axes. **B.3:** Combined movement task, commonly encountered in VEs entailing MR technologies and robotic teleoperation (Triantafyllidis and Li, 2021a; Triantafyllidis et al., 2021). Necessitates the most complex motor movements due to the large number of entailed spatial variables.

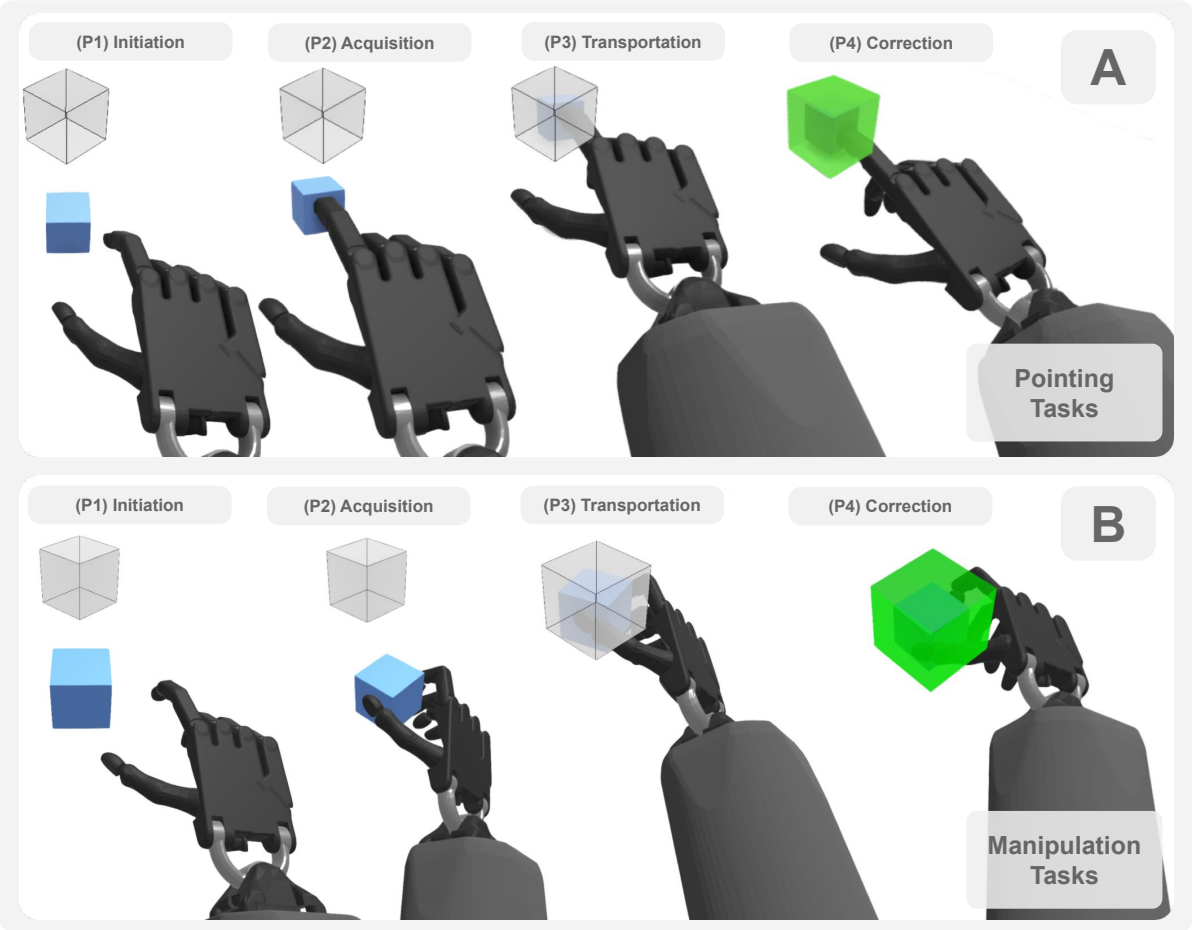

Figure 4.4: **Overview of different motor tasks with different types of motor interactions.** The figure depicts the two types of interactions studied, entailing pointing and manipulation across four phases (P1-P4). **Figure A:** Pointing tasks. **Figure B:** Manipulation tasks entailing grasping. **Note:** The main difference between A and B lies in phase P2, where the object is either grasped or attached to the hand depending on the type of interaction.

Table 4.3: **The definition of task success rates for different concepts of motor movement types.** The table represents mathematically the task successes (1) or errors (0) for the assessment of translational, rotational and fully combined 3D motor tasks.

Translational Tasks	**Rotational Tasks**	**Combined Tasks**
$T = \begin{cases} 1, & A \leqslant \frac{W}{2} \\ 0, & A > \frac{W}{2} \end{cases}$	$R = \begin{cases} 1, & \alpha_i \leqslant \omega_i \\ 0, & \alpha_i > \omega_i \end{cases}, i : \mathbb{R}^3$	$C = T \cdot R$
Notations for Translational Tasks: Variable A represents the distance between two 3D points, hence $A = d(W_{x,y,z}, F_{x,y,z}) = \sqrt{\sum_{i=1}^{3} (x_i - y_i)^2}$.		
Notations for Rotational Tasks: Term $\alpha_i = \lvert W_{\alpha_i} - F_{\alpha_i} \rvert$. W_{α_i} and F_{α_i} represent the rotation of the target and the object in all axes (i) respectively. $\alpha_i' = \alpha_i$.		
Notations for Combined Tasks: Both translational and rotational requirements need to be met to be classified as a success e.g. $T = 1$ and $R = 1$.		

Task Design – In terms of task design and more specifically for pointing tasks, the index pad finger of the virtual hand was used, to virtually "attach" the cube to the participant's hand to allow users to move their hand and the cube simultaneously, to the target location. The object would attach during collision via the intersection between the index finger of the participant and the cube. Moreover, during the attachment, the cube would match the current position of the pad index finger but retain its original orientation. Instead, for the manipulation types of tasks, realistic physics and friction forces were simulated in the simulation environment. During such tasks, participants had to physically grasp and transport the object to the target location by simulated contact forces. Gravity ($g = 9.8\ m/s^2$) was enabled for manipulation and disabled for pointing. Moreover, a table was simulated by a collision plane for the manipulation types of task to render the retrieval of the cube object possible. Figure 4.4.A and Figure 4.4.B visually depict pointing and manipulation types of interactions respectively.

4.3 Experiments and Evaluation

This section presents the outline of the experimental design and the subsequent evaluation approach of the acquired data. Details are provided on the recruitment process of participants and the selection criteria set, the acclimatisation to the simulation environment as well as the subsequent statistical analysis used for the recorded data.

4.3.1 Pre-Exposure and Assessment of Experiments

For this work, a total number of twenty participants ($N = 20$) were recruited to take part in the experiments. From the total number of participants, 4 were females and 16 males, with ages ranging from 19 to 46 ($\mu = 27.35$, $\sigma = 5.43$). For the recruitment, certain selection criteria were put forth, namely, all participants (i) were right-handed, (ii) stated that they had healthy hand control with (iii) normal to corrected vision and (iv) were familiar with either video games or previous VR experiences. Participants had to meet all four aforementioned criteria to be included in the experiment. Prior to the experiments, participants were asked to find a balance between minimising errors and selecting the targets as quickly as they could during target selection and placing of the motor tasks. These decisions were made to retain consistency during recruitment and increase the generalisation of the results of this work. Consult Figure 4.1, Figure 4.3 and Figure 4.4 for a general overview of the motor assessment tasks.

Prior to the commencement of the experiment, all participants were briefed about the experiment and the aim of this research project. All participants voluntarily took part in the study by providing their formal written consent. Since the experiments entailed the use of VR hardware, the individual IPD for each participant was measured for the VRHMD. Moreover, acclimatisation to the simulation environment was allowed for all participants prior to the experiment, via a set of 96 training exercises covering both pointing and manipulation types of task interactions with the addition of translational and rotational movements. These training exercises were specific to the acclimatisation procedure and independent of those in the experiment to disallow task adaption to the actual motor tasks. Furthermore, the order of all experiments was randomised for each participant to counterbalance potential acclimatisation or task adaption. A grand total of 39,040 trials were recorded, of which only successful instances were included for the subsequent evaluation.

4.3.2 Statistical Analysis and Procedure

Similarly to Chapter 3 and more specifically Section 3.4.4, a similar statistical analysis and procedure were followed for the investigation of whether spatial variables in increasing complexity motor tasks had a significant effect towards completion time. This was considered critical to ensure a comprehensive understanding of which spatial variables would potentially be important for the derivation of a novel metric extension based on Fitts'

Law. This, in turn, would allow for the measurement and quantification of upper-body biomechanical movements stemming from the intricate sensory-motor system in humans.

The investigated spatial variables were tested for significance (95% CI) via the use of a RM-ANOVA. A Shapiro-Wilk Test was used to verify the normality of the data prior to the RM-ANOVA and in cases of violation, i.e. non-parametric data, an ART (Wobbrock et al., 2011a) was used to allow the use of the RM-ANOVA on the ranked data. Maulchy's Test of sphericity was used to test sphericity of the data from the RM-ANOVA and in cases of violation, a GGC was used to account for the violation and correct the degrees of freedom assuming $\epsilon < .75$. A step-wise linear regression determined the effect of each variable for conceiving a new metric equation and the effect on the predictability of MT which allowed, in turn, the author to fully understand which ones were considered crucial for human motor movement. The independent spatial variables stemming from the experiments were either fitted using the criteria of the probability of $F < .05$ to enter, or $F > .10$ to remove. Finally, linear regression analysis was used (r^2) to analyse and compare the proposed derived model with existing work to infer which model best explained the data stemming from the motor task experiments. Similarly to Chapter 3, all reported significance levels in this chapter are: $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$, and n.s (or $p > .05$) standing for "not significant".

4.4 Results

In this section, all results stemming from four in total increasing spatial assessment experiment tasks, hereinafter referred to as **Motor Tasks**, are presented. At each stage of those motor tasks, the data was interpreted and analysed for both pointing and manipulation types of interaction, with more details regarding their definitions provided in Section 4.2.3.

In Figure 4.5, the relationship between all included spatial variables and MT is visually depicted. Moreover, in Table 4.2, the variations of those investigated variables in each of the Motor Tasks tested are detailed. In Table 4.4 and Table 4.5 summarise the statistical results and significance levels across all motor tasks for pointing and manipulation types of interactions respectively. Finally, in Table 4.6 and Figure 4.6, all motor performance models are summarised and compared with their respective r^2 values, in addition to the data points and line fittings, respectively.

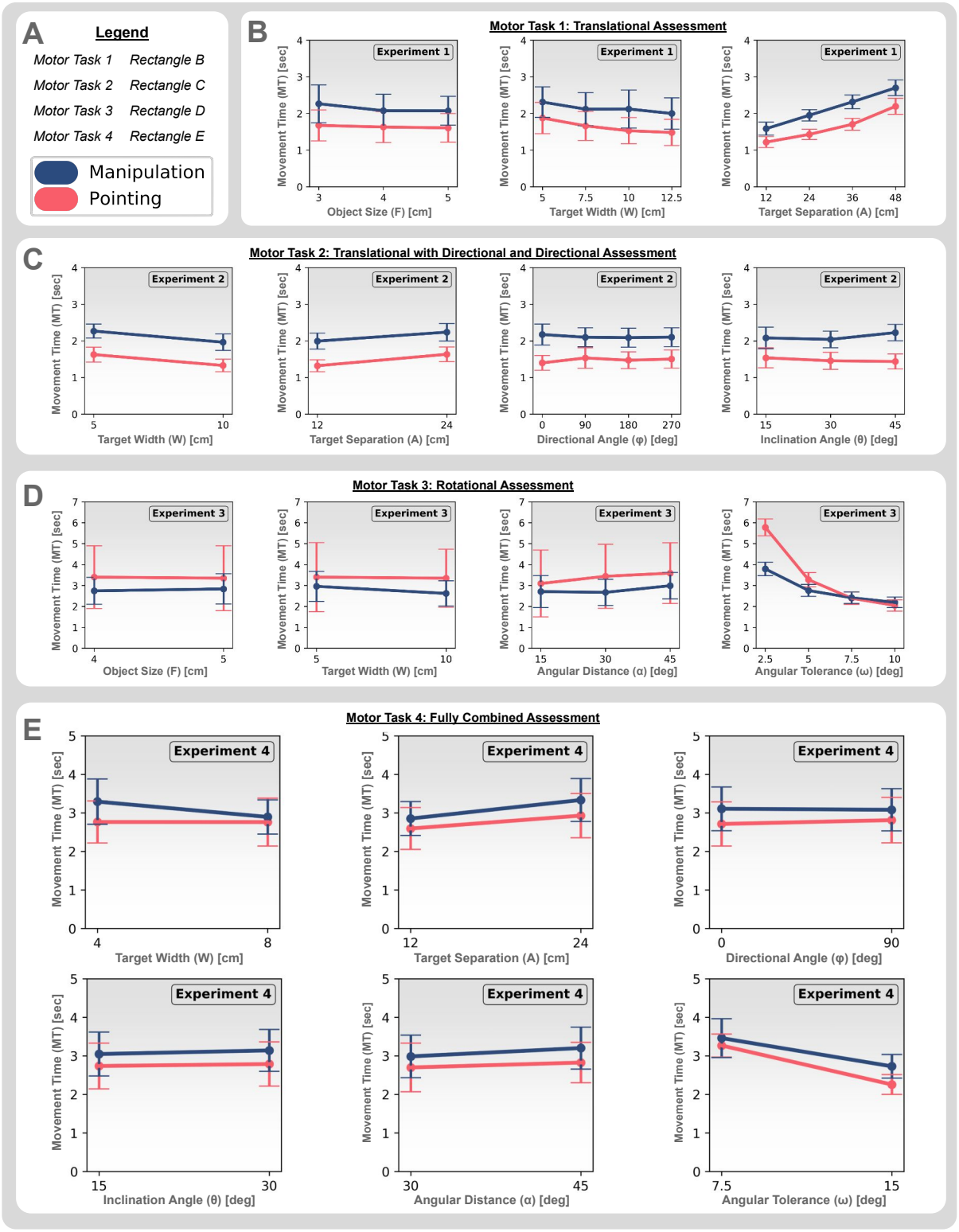

Figure 4.5: **Results of the effects of all spatial variables.** The figure depicts the effects of all spatial variables on *MT* across all four motor tasks for both pointing (red) and manipulation (blue) types of interactions. Bars represent the standard deviation. **Figure A:** Legend. **Figure B:** Motor Task 1, entailing the assessment of translational motions. **Figure C:** Motor Task 2, entailing the assessment of translational motions with the inclusion of directions and inclinations. **Figure D:** Motor Task 3, entailing the assessment of rotational motions. **Figure E:** Motor Task 4, the final, fully combined movement task entailing the assessment of both translational and rotational motions with directions and inclinations.

4.4.1 Motor Task 1 – Translational Assessment

The initial aim of this research was to approximate to the closest possible extent Fitts' original experiment, except in 3D space. Consequently, in this first motor assessment task, the effects of purely translational tasks were investigated, in particular on the variables of object size F, target width W and target separation A. Hence, movements were along one single line only, without the addition of either directions or inclinations to keep the first 3D task as simple as possible.

Design of Motor Assessment Task

In the first experiment, a 3x4x4 within-subjects design was used with the independent variables being: three object sizes (F = 3, 4 and 5cm), four target sizes (W = 5, 7.5, 10, 12.5cm) and four target separations (A = 12, 24, 36, 48cm). The dependent variable was movement time MT measured in seconds. Among the 48 types of tasks, each had 5 repetitions for both pointing and manipulation types of interaction, accounting for a total of 480 trials. With 20 participants, a grand total of 9,600 trials were recorded.

Results of Motor Assessment for Pointing Tasks

The data was not normally distributed as yielded by a Shapiro-Wilk Test ($p = .029$), and thus an ART was used prior to the RM-ANOVA to allow the subsequent analysis on the ranked data. The mean MT for pointing tasks was 1.63 ± 0.40s. Four trials out of 4,800 were excluded from the analysis due to errors, accounting for a total of 4,796 valid trials. From the results, it was inferred that an increase in object size F and target width W significantly in turn decreased MT, with significance values equalling $p < .01$ and $p < .001$ respectively. In line with F and W, the target separation A also significantly influenced MT, yet at a much greater extent with an almost linearly increasing correlation ($p < .001$).

Results of Motor Assessment for Manipulation Tasks

The data for the manipulation types of tasks was normally distributed as yielded by a Shapiro-Wilk Test ($p = .229$), with the mean MT equalling 2.13 ± 0.45s. In total 44 trials out of 4,800, i.e. approximately 0.9% of the data, were excluded from the analysis due to errors, leaving 4,756 valid data. MT significantly decreased as the object size F and target width W increased, ($p < .001$) and ($p < .001$) respectively. In line with the pointing results, the target separation A did once again influence MT the most, with a

similar linearly increasing correlation MT as yielded by a RM-ANOVA ($p < .001$).

Remarks

The spatial variables and their effect on MT are visually shown in Figure 4.5.B. From the above analysis, it can be inferred that all included spatial variables as seen in Fitts' original formulation and that of Hoffmann's, had a significant effect on MT for both pointing and manipulation types of interaction. More specifically, fitting the models of Fitts', MacKenzie's and Welford's on the data revealed high fitting results (r^2) for both pointing and manipulation, with the latter showing slightly less fitting. The only model that showed significantly better fitting for manipulation, at an approximately 10% increase from the rest was Hoffmann's formulation, attributed to the incorporation of the object size F in the model. The models of Murata and Iwase' and Cha and Myung's were excluded from this analysis as they yielded the same results as their based extensions since directions and inclinations were not assessed until Motor Task 2. Regression results with r^2 fittings are shown in Figure 4.6.

4.4.2 Motor Task 2 – Translational with Directional and Inclinational Assessment

This motor task is the extension of the previous Motor Task 1 with the exception of including both directional and inclinational spatial arrangements. The motivation for adding these spatial arrangements stemmed from the literature review indicating significant effects of directions and inclinations towards completion time (Cha and Myung, 2013; Murata and Iwase, 2001). Object size F was excluded from this experiment to effectively limit this study to the bare minimum spatial variables of target width (W) and target separation (A). Instead, emphasis was given primarily on directional (ϕ) and inclinational θ arrangements to understand their potential importance towards motor task performance in 3D space.

Design of Motor Assessment Task

In this motor task experiment, a 2x2x4x3 within-subjects design was used. The independent variables were: two target sizes ($W = 5$ and 10cm), two target separations ($A = 12$ and 24cm), four direction angles ($\phi = 0°$, $90°$, $180°$ and $270°$) and three inclination angles ($\theta = 15°$, $30°$ and $45°$). $F = 5$cm. The dependent variable was movement time MT. A

grand total of 9,600 trials were recorded.

Results of Motor Assessment for Pointing Tasks

The data was normally distributed ($p = .077$), with the mean MT being 1.47 ± 0.23s. A total of 3 trials out of 4,800 were excluded from the analysis due to errors, leaving 4,797 valid ones. Spatial variables once again had a significant effect towards MT as with Motor Task 1, with increasing target width W values significantly decreasing MT ($p < .001$), while higher values of target separation A significantly increased MT ($p < .001$). Directional angles ϕ significantly influenced MT as well ($p < .05$), with a slight sinusoidal relationship with MT. More specifically, the results revealed that front-to-backward movements (0°, 180°) take slightly longer time to complete than left-to-right ones (90°, 270°). These results appear to be in line with Barrera Machuca and Stuerzlinger (2019). Lastly, higher degrees of inclinations θ translated in longer overall completion times MT, ($p < .05$), which are also in line with the findings of Cha and Myung (2013).

Results of Motor Assessment for Manipulation Tasks

The data was normally distributed ($p = .083$) as with the pointing types of interaction. The mean MT for manipulation tasks was 2.11 ± 0.25s. A total of 60 trials out of 4,800, 1.25% of the data, were excluded from the analysis due to error, leaving 4,740 valid ones. Similarly, with the pointing tasks, manipulation tasks with directional angles appear to still be dependent on the fundamental spatial variables of target width W and target separation A, both of which significantly affected MT. More specifically, increased values of target width W significantly decreased MT ($p < .001$), while increased values of target separation A significantly increased MT ($p < .001$). Contrary to pointing tasks, directional angles ϕ did not have an apparent effect on MT as yielded by the RM-ANOVA ($p = .476$). In line with the pointing tasks, increased incline angles θ did significantly account for longer overall MT ($p < .05$), as with Cha and Myung (2013).

Remarks

The spatial variables and their effect on MT are visually shown in Figure 4.5.C. From the above results it can be inferred that all included spatial variables had significant effects on MT, except directional angles ϕ when presented with physical manipulation. Contrary to Motor Task 1, it was observed that Fitts', MacKenzie's and Welford's formulation had

a relatively higher fitting both in pointing and manipulation than that of Hoffmann's model by approximately 5%. These results are furthermore confirmed when fitting Cha and Myung's model, which is essentially based on Hoffmann's, having a very similar fitting for both pointing and manipulation. Lastly, only Murata and Iwase's directional model presented a marginally higher fitting than the rest in both pointing and manipulation tasks, underlining to an extent the importance of spatial arrangements. Regression results with r^2 fittings are shown in Figure 4.6.

4.4.3 Motor Task 3 – Rotational Assessment

All experiments assessing motor movement to this point only included and investigated translational tasks. However, as commonly seen in simulation training, MR and most importantly in robotic teleoperation, rotation is an inseparable and fundamental part (Triantafyllidis and Li, 2021a; Kulik et al., 2020; Stoelen and Akin, 2010). Consequently, in this particular motor task, translation is omitted to fully focus on rotational variables. To achieve this, both the object (F) and the target (A) will be overlapping in the centre, with only their rotation differing. With the above design decision, this task renders itself an almost natural equivalent of Fitts' original experiment, only for rotation.

Design of Motor Assessment Task

The study used a 2x2x3x4 within-subjects design. The independent variables were: two object sizes ($F = 4$ and 5cm), two target sizes ($W = 5$ and 10cm), three target rotations ($\alpha = 15°$, $30°$ and $45°$) and four rotational tolerances ($\omega = 2.5°$, $5°$, $7.5°$ and $10°$). The dependent variable was movement time MT. 9,600 trials were recorded.

Results of Motor Assessment for Pointing Tasks

The data significantly deviated from normality ($p < .001$) as yielded by a Shapiro-Wilk Test. Consequently, an ART was applied to allow the analysis of the ranked data via the RM-ANOVA. The mean MT for pointing tasks was 3.37 ± 1.51s. A total of 225 trials out of 4,800, 4.6% of the data, were excluded from the analysis due to error, leaving 4,575 valid ones. Contrary to the findings up until this point, an increase in both object size F and target width W showed a slight decrease towards MT, but not at a significant level for either variable, with the RM-ANOVA yielding $p = .226$ and $p = .590$ respectively. However, it was observed that a significant correlation between the rotational separation

α and MT was revealed ($p < .001$), with higher values of α corresponding to higher MT. The largest effect was observed with the rotational tolerance ω, revealing an almost inverse exponential relationship with MT ($p < .001$), with lower values of ω significantly increasing MT. From these results, it can be inferred that the effect of ω was so significant that it had almost the same inverse effect on MT as the equivalent of W and F for translational tasks, though at an apparent much higher degree.

Results of Motor Assessment for Manipulation Tasks

The data violated normality ($p < .01$) as shown by a Shapiro-Wilk Test. Hence, an ART was used to allow the analysis of the ranked data via the RM-ANOVA. The mean MT for manipulation tasks was 2.79 $\pm$ 0.67s. A total of 64 trials out of 4,800, 1.3% of the data, were excluded due to error, leaving 4,736 valid data. Similarly to pointing, no significant effect was observed on MT with the object size F ($p = .286$). Interestingly enough and in contrast to pointing types of tasks, an increase in target width W did significantly decrease MT ($p < .001$). In line with pointing, increasing rotational separation α also increased MT significantly ($p < .001$) when manipulating objects. Finally, as with pointing, increasing rotational tolerance ω significantly decreased MT ($p < .001$) in a non-linear fashion during manipulation. It can hence be inferred that with the exception of target width W, the same relationship of the tested variables on MT for both pointing and manipulation to this point for the rotational tasks is observed. As shown in both interaction types of cases, rotational tolerance ω predominately affected MT in a purely rotational setting.

Remarks

The spatial variables and their effect on MT are visually shown in Figure 4.5.D. For this motor task assessment, all existing equations from the state-of-the-art are fitted and adapted to accommodate the rotational nature of this experiment, as detailed in Table 4.6. In short, Fitts' original formulation defined as $ID_t = log_2(\frac{2A}{W})$ was adjusted as $ID_r = log_2(\frac{2\alpha}{\omega})$, to allow the model to be evaluated for the rotational nature of this experiment. The other model extensions were adjusted similarly. This allowed, in turn, to conduct and apply these methods to the rotational nature of this experiment while also investigating any significant effects for comparison. This led this research to further extend the work of Stoelen and Akin (2010) and cover, to the best of the author's knowledge, all known extensions to this date. Overall, regarding variables influencing MT, it can be inferred that rotational variables α and ω were significant in both cases, although

some inconsistencies were observed between pointing and manipulation. Contrary to Motor Tasks 1 & 2, none of the adjusted models fitted this rotational task well, where all formulations were below $r^2 \leq 0.680$ for both pointing and manipulation types of interaction, showing their current limitations in 3D space when assessing rotation. Regression results with r^2 fittings are shown in Figure 4.6. To more comprehensively examine the impact of spatial variables, particularly the rotational terms and their influence on motor completion times, Motor Task 4 was designed to integrate all previously examined spatial parameters within a single setting. This design enabled a full investigation of the underlying factors of human movement in 3D space.

4.4.4 Motor Task 4 – Fully Combined Assessment

For the final motor task experiment, all possible variations from Motor Task 1, 2 and 3 were combined. Consequently, this experiment represented a fully combined movement task. In this motor assessment task, translational and rotational task variations were fully combined, rendering it a full 3D with 6DoFs task. Due to the plethora of different spatial variables involved in this experiment, the number of variations was limited to keep the experiment short, which would have otherwise resulted in prohibitively long duration for the participants. Consequently, the directional variations of ϕ were restricted to the front and right side directions (0°, 90°), while still effectively investigating view and lateral directional influences, which have been shown to differ in difficulty and influence MT according to the works of Barrera Machuca and Stuerzlinger (2019). Moreover, the object size was fixed at $F = 4$cm. These design decisions resulted in the already exhaustive final study being within an acceptable duration for the participants.

Design of Motor Assessment Task

A 2^6 within-subjects design combination was used for the final motor assessment task. The independent variables were: two target sizes ($W = 4$ and 8cm), two target separations ($A = 12$ and 24cm), two direction angles ($\phi = 0°$ and 90°), two inclination angles ($\theta = 15°$ and 30°), two target rotations ($\alpha = 30°$ and 45°) and two rotational tolerances ($\omega = 7.5°$ and 15°). The dependent variable was movement time MT. Consequently, this represented a total of 64 tasks, with 4 repetitions, each for pointing and manipulation, for a total of 512 trials per participant. This resulted in a grand total of 10,240 trials.

Results of Motor Assessment for Pointing Tasks

The data deviated from normality as yielded by the Shapiro-Wilk test ($p = .016$) and thus an ART was performed to allow the subsequent analysis via the RM-ANOVA on the ranked data. The mean MT for pointing tasks was 2.71 $\pm$ 0.55s. A total of 120 error trials out of 5,120, 2.5% of the data, were excluded from the analysis leaving 4,992 valid data points. Sphericity was met in all cases since all the tested variations were two levels, unlike the previous motor task assessments to this point. Foremost, there was no significant observed effect of target width W on MT, ($p > .05$). However, target separation A, consistent with Motor Tasks 1 and 2 concerned with translational movements with the addition of directional and inclinational angles respectively, significantly increased MT with higher separation values ($p < .001$). Interestingly enough, neither directions ϕ ($p = .141$) nor inclinations θ ($p = .200$) had any significant effect on MT, even though marginal effects were observed. Finally, similarly to Motor Task 3, rotational separation α significantly affected MT ($p < .01$) with higher values significantly increasing MT, while ω had the highest influence on MT ($p < .001$), with lower rotational tolerance values being inversely proportional to the difficulty of the 3D task, leading to significantly higher MT.

Results of Motor Assessment for Manipulation Tasks

Unlike pointing tasks, the data from the manipulation tasks was normally distributed as yielded by a Shapiro-Wilk Test ($p = .091$). The mean MT for manipulation tasks was 3.10 $\pm$ 0.59s. A total of 82 trials out of 5,120, $\approx$ 1.7% of the data, were excluded from the analysis due to error leaving a total of 5,032 valid data points. Sphericity was met in all cases as with pointing tasks. Contrary to the results observed pointing task, target width W had a significant effect on MT ($p < .001$) with higher dimensions leading to decreased completion times. Target separation A, consistent with all the findings to this point, significantly increased MT with higher separation values ($p < .001$). Consistent with the pointing tasks of this experiment, directional angles ϕ, as well as inclination angles θ, did not affect MT with significance values equalling $p = .099$ and $p = .227$ respectively. As expected to this point, rotational separation α did significantly affect the increase of MT as the degrees of separation rose ($p < .01$). Rotational tolerance ω did once again account for the highest influence on MT, being consistent with the pointing equivalent and Motor Task 3 concerned with rotation ($p < .001$). More specifically, lower values of ω essentially constituting higher rotational accuracy, led to significantly higher difficulty (i.e. MT).

Remarks

For this experiment, similarly to Motor Task 3, the IDs of translation and rotation were added for all existing approaches to allow for the comparison of the methods taking into account both existing types of motion. More specifically, rotation was adjusted as described in Motor Task 3 and in Table 4.6. It is worthwhile to point out that these models are in principle incompatible with combined movements entailing translation and rotation. Consequently, it was deemed important to adjust these methods and add the two IDs for both translation and rotation to explore the potential of summing these two separate concepts of motion and to conduct a fair comparison. For instance, the combined ID for Fitts' original model would be transformed into: $ID_t = log_2(\frac{2A}{W})$ for translation, plus $ID_r = log_2(\frac{2\alpha}{\omega})$ for rotation. The other formulations used to this point followed the same fashion and adjustment to allow for their comparison and model fitting.

The spatial variables and their effect on MT are visually shown in Figure 4.5.E. Overall, from the results it can be inferred that neither directional ϕ nor inclination θ angles had any significant effect on MT ($p > 0.05$). This was consistent for both pointing and manipulation. For combined movements, adding both IDs of translation and rotation, Fitts's, MacKenzie's, Welford's and even Murata and Iwase's directional formulation proved insufficient when fitted for pointing (all yielding $r^2 < 0.5$). Interestingly enough, those models yielded slightly better results for manipulation but still to a lower degree ($r^2 < 0.77$). On the other hand, Hoffmann's model had a significantly higher fitting towards pointing with an approximately 10% increase likely attributed to the incorporation of the object size F in the equation. Consistent with Motor task 1, this suggests the necessity of adding the object size F in a model and the value of Hoffmann's formulation. The model of Cha and Myung's had the highest fitting on the data but only marginally over the rest, which was primarily due to the incorporation of object size F, as it is greatly based on Hoffmann's formulation. Regression results with r^2 fittings are shown in Figure 4.6.

In the next section, the derivation of a novel model is presented. The derivation of this model is made possible based on all of the findings and effects of the studied spatial variables from all motor tasks to this point. Given the state-of-the-art model extensions of Fitts' law and their characteristics, a novel metric is derived, collectively combining the entailed spatial variables as encountered in full 3D motions with 6DoFs.

Table 4.4: **Statistical results of all variables and their influence towards movement time for pointing tasks.** The table details all spatially investigated variables and their effect on MT across all experiments for pointing types of interactions. The percentages from r^2 (%) indicate the weight or contribution of each variable towards MT and were obtained via step-wise linear regression.

Statistical Results of Spatial Variables Across Motor Tasks for Pointing							
	Interaction Type: Pointing						
	Sphericity Test			*RM-ANOVA Test*			
Variables	χ^2	p	ϵ	F-Test	η_p^2	p	r^2 (%)
Motor Task 1							
F	$\chi^2(2)$=2.468	n.s	n/a	F(2,30)=9.441	0.386	**	0.5%
W	$\chi^2(5)$=6.992	n.s	n/a	F(3,33)=102.522	0.903	***	10.2%
A	$\chi^2(5)$=8.360	n.s	n/a	F(3,33)=602.976	0.982	***	80.4%
Motor Task 2							
W	n/a	n/a	n/a	F(1,23)=287.752	0.925	***	40.0%
A	n/a	n/a	n/a	F(1,23)=190.993	0.893	***	43.7%
ϕ	$\chi^2(5)$=12.609	*	0.675	F(2.025,22.270)=6.918	0.386	*	3.0%
θ	$\chi^2(2)$=4.346	n.s	n/a	F(2,30)=7.160	0.323	*	1.3%
Motor Task 3							
F	n/a	n/a	n/a	F(1,23)=1.551	0.063	n.s	0.0%
W	n/a	n/a	n/a	F(1,23)=0.299	0.013	n.s	0.0%
α	$\chi^2(2)$=0.781	n.s	n/a	F(2,30)=23.639	0.612	***	1.8%
ω	$\chi^2(5)$=11.386	*	0.639	F(1.918,21.093)=184.323	0.944	***	81.8%
Motor Task 4							
W	n/a	n/a	n/a	F(1,31)<0.001	0.0	n.s	0.0%
A	n/a	n/a	n/a	F(1,31)=51.296	0.623	***	8.5%
ϕ	n/a	n/a	n/a	F(1,31)=2.278	0.680	n.s	0.7%
θ	n/a	n/a	n/a	F(1,31)=1.718	0.053	n.s	0.2%
α	n/a	n/a	n/a	F(1,31)=10.548	0.254	**	1.2%
ω	n/a	n/a	n/a	F(1,31)=394.944	0.927	***	77.3%

Table 4.5: **Statistical results of all variables and their influence towards movement time for manipulation tasks.** The table details all spatially investigated variables and their effect on MT across all experiments for manipulation types of interactions. The percentages from r^2 (%) indicate the weight or contribution of each variable towards MT and were obtained via step-wise linear regression.

Statistical Results of Spatial Variables Across Motor Tasks for Manipulation							
	Interaction Type: Manipulation						
	Sphericity Test			*RM-ANOVA Test*			
Variables	χ^2	p	ϵ	F-Test	η_p^2	p	r^2 (%)
Motor Task 1							
F	$\chi^2(2)$=3.314	n.s	n/a	F(2,30)=14.468	0.491	***	3.1%
W	$\chi^2(5)$=5.409	n.s	n/a	F(3,33)=19.162	0.635	***	5.3%
A	$\chi^2(5)$=5.968	n.s	n/a	F(3,33)=197.465	0.947	***	84.7%
Motor Task 2							
W	n/a	n/a	n/a	F(1,23)=51.431	0.691	***	35.6%
A	n/a	n/a	n/a	F(1,23)=33.986	0.596	***	23.0%
ϕ	$\chi^2(5)$=3.722	n.s	n/a	F(3,33)=0.851	0.072	n.s	1.0%
θ	$\chi^2(2)$=2.210	n.s	n/a	F(2,30)=5.535	0.270	*	5.4%
Motor Task 3							
F	n/a	n/a	n/a	F(1,23)=1.195	0.049	n.s	0.5%
W	n/a	n/a	n/a	F(1,23)=40.805	0.640	***	6.4%
α	$\chi^2(2)$=3.589	n.s	n/a	F(2,30)=12.915	0.463	***	2.9%
ω	$\chi^2(5)$=4.454	n.s	n/a	F(3,33)=89.481	0.891	***	73.3%
Motor Task 4							
W	n/a	n/a	n/a	F(1,31)=58.815	0.655	***	13.1%
A	n/a	n/a	n/a	F(1,31)=48.664	0.611	***	19.2%
ϕ	n/a	n/a	n/a	F(1,31)=0.205	0.007	n.s	0.1%
θ	n/a	n/a	n/a	F(1,31)=2.990	0.088	n.s	0.7%
α	n/a	n/a	n/a	F(1,31)=10.193	0.247	**	4.0%
ω	n/a	n/a	n/a	F(1,31)=122.753	0.798	***	44.3%

4.5 Metric Derivation

As identified in the related work, there is a plethora of work making use of Fitts' law with a multitude of studies deriving extensions from it (Triantafyllidis and Li, 2021a). However, a clear definition of the associated spatial variables within the formulation is mostly missing (Drewes, 2010). In particular, the distance of the cursor, or object when referring to manipulation tasks, to the target location, defined as A, can either be formulated and defined as the distance of the cursor to the target centre (A_{cc}), the cursor's centre to the target edge (A_{ce}) or the cursor edge to the target edge (A_{ee}). While these definitions of distances may not initially seem to be of importance, they do in actuality influence the formula, as:

$$A_n \begin{cases} A_1 = A_{cc}, \\ A_2 = A_{ec} + \frac{W}{2}, \\ A_3 = A_{ee} + \frac{W+F}{2}, \end{cases} \qquad d_n = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \tag{4.2}$$

As it can be inferred from the aforementioned, a total of three distinct target separations can be defined as A_n with $A_{cc} > A_{ec} > A_{ee}$. The Euclidean distance between the cursor and the target, i.e. the two points in 3D space, is represented as d_n with $n = 3$.

In every stage where increasing levels of motor task complexity were assessed, as presented gradually in Section 4.4, it was shown that 3D translational movement followed Fitts' formulation fairly closely. This was a surprising finding due to the Fitts' formulation originally being applicable in 2D space. In particular, throughout all Motor Tasks entailing 3D translational movements, the terms target separation A and target width W were found to be inseparable parts. However, it was also observed that another variable that proved to have a determining effect during the initial assessment of translational tasks was the object size F. Interestingly enough, the object size (F) is part of Hoffmann's model (Hoffmann, 1995) but not of Fitts' (Fitts, 1954). Consequently, it can be inferred from Equation 4.2 and the observations made to this point, that the most appropriate definition of effective target separation would be A_3, which would include both W and F. By incorporating this definition into Fitts' original model (ID) shown in Equation 2.1, the following equation would be derived:

$$ID_t = log_2 \left(\frac{2A}{F+W} + 1 \right). \tag{4.3}$$

Following the higher motor complexity assessment tasks, it was inferred that for rotational movements, Figure 4.5, Table 4.4 and Table 4.5 revealed that the rotational separation

between the cursor (or object for manipulation) and the target (α) followed an almost linear relationship with MT. However, the rotational tolerance (ω) followed an evident non-linear relationship. Following further analysis, it was observed that ω follows an almost inverse exponential decay and a $\frac{1}{x^2}$ would be the most appropriate model fitting that could explain the data from the experiments. Nevertheless, more data points would be needed in future work to shed additional light on this discrepancy and increase the understanding of future findings of the importance of the included rotational variables.

From the results at hand, the subsequent analysis and the effect of included spatial variables towards motor completion times, the author of this thesis formulates an improved and more suitable 3D extension of Fitts' law as follows:

$$\text{Final Model: } \begin{cases} MT &= a + b\ [c \cdot ID_t + d \cdot ID_r], \\ ID_t &= log_2\left(\frac{2A}{F+W} + 1\right), \\ ID_r &= log_2\left(\frac{2\alpha}{\omega^2} + 1\right), \end{cases} \tag{4.4}$$

whereby the variables a, b, c and d are constants determined through regression. As it may be referred, Fitts' original approach is retained, and hence constants a and b represent the y-intercept and slope respectively. Moreover, the constants c and d represent the contribution of translation (ID_t) and rotation (ID_r) respectively, as these are two separate concepts of motion as evidenced from the analysed results.

It is worthwhile to point out that the constant b is superfluous in this case since it multiplies constants c and d. Nevertheless, retaining the b constant allows the resemblance and familiarity of the derived model to that of Fitts' original law as detailed in Equation 2.1.

The derived model follows the work of Stoelen and Akin (2010) closely. More specifically, this work and the subsequent findings, support the observations of Stoelen and Akin (2010) and in particular the feasibility of adding the IDs of both translational and rotational motor tasks into one and unified combined task ID, however, with one important difference. In their work, Stoelen and Akin (2010) found that both ID_t and ID_r had apparent equal contribution towards MT as both terms were controlled by a single constant b. This led to the formulation of: $MT = a + b \cdot [ID_t + ID_r]$. Consequently, their implication was that both indices of difficulty have an identical contribution (i.e. 50%) towards MT, which from the results of this work appears to be imprecise.

From the results at hand and the subsequent analysis of the increasing complexity of motor assessment tasks in full 3D space, it is shown that rotational motor movements

almost predominately affected MT. In particular, this was primarily attributed to high rotational tolerance (or accuracy) as seen with $\omega = 2.5°$ in Motor Task 3 – Rotational Assessment. These findings were furthermore confirmed in combined translational and rotational motions in Motor Task 4. For very small values of rotational accuracy (ω), as aforementioned, an almost inverse exponential decay was observed with MT instead of a linear one as implied by Stoelen and Akin (2010). **Hence, translation and rotation should be perceived as separate terms, and each of those concepts of motion should be quantified by its own constants.** This would mitigate the limitation of equally weighing translational and rotational difficulties as with Stoelen and Akin (2010), as this is not supported by the results of this work and in fact, rotational motion appears to be significantly more demanding and by extent requiring longer completion times.

In the following section (Section 4.6), the results stemming from this work are analysed in more in-depth. Most importantly, the derived model formulation from this work is compared with other state-of-the-art model extensions of Fitts' law.

4.6 Discussion on Quantifying Motor Movements

In this section, the results and implications of this work regarding capturing and assessing human movement stemming from the intricate sensory-motor system will be discussed. As presented in the literature review in earlier parts of this thesis, being able to combine multiple types of metrics in one formulation is highly beneficial in research as it allows for the comparability of results amongst different works. Fitts' Law represented itself as a promising formulation for this endeavour, capable of quantifying human speed based on motor-related spatial complexities, essentially combining both time and spatial-based evaluation metrics in one formulation and being highly prominent amongst existing work (Triantafyllidis and Li, 2021a).

4.6.1 Implications of Results

From the four in total designed motor assessment tasks entailing upper-body movements in a simulated teleoperation setting with the use VR technologies and adding increasing spatial complexity, the most widely used extensions of Fitt's law were compared and tested. The applicability of each model formulation was evaluated within those experiments entailing two types of interactions; pointing and manipulation. These types of interactions

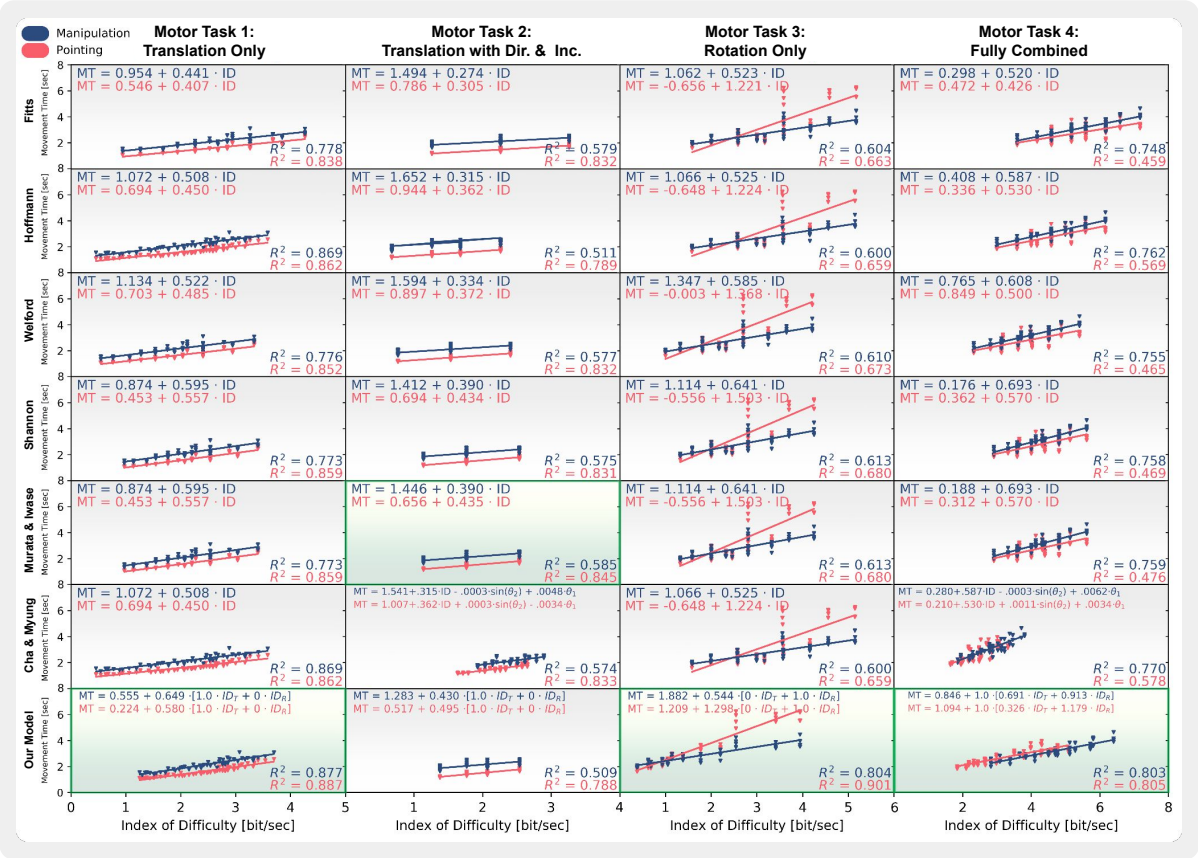

Figure 4.6: **Regression plots of all performance models across all motor tasks.** The figure depicts the regression results of all models across all four motor tasks with increasing spatial complexity. Both pointing (red) and manipulation (blue) types of interactions are detailed for each model. Line equations and r^2 are detailed for each model. Green boxes represent the best model for the column category which in turn corresponds to the particular motor task studied.

are the most widely used in collaborative VEs, robotic teleoperation and more broadly human ergonomics cases (Barrera Machuca and Stuerzlinger, 2019; Triantafyllidis et al., 2020; O'Hara, 1987).

From these results, it can be inferred that in the most basic form of 3D motor movement in a purely translational setting (Motor Task 1), Fitts' law and its extensions, are sufficiently adequate for explaining both target pointing and manipulation, limited to a one-directional line only. However, during increased spatial complexity entailing directional and inclination angles around the participants (Motor Task 2), the current model formulations showed reduced performance at predicting the results of the experiments, which dropped slightly in pointing tasks but significantly more in manipulation types of tasks.

As it can be inferred from the related work, assessing purely translational movement is limited, as most motor movements seen by humans entail rotational variations as well (Stoelen and Akin, 2010; Triantafyllidis and Li, 2021a). Even though existing model extensions of Fitts' law did attempt to model rotation, their findings are still limited in 2D space (Kondraske, 1994; Kulik et al., 2020; Stoelen and Akin, 2010). With the exception of Kulik et al. (2020) and Stoelen and Akin (2010) who did evaluate combined movements, these were nonetheless still limited to 2D space. Indeed, after further reviewing related work, it was inferred, to the best of the author's knowledge, that combining rotational and translational movements under one setting in 3D remains largely unexplored.

From the above, it can be confirmed that even combining translational and rotational variations by equally weighing their difficulty and influence towards MT is limited. More specifically, all model formulations were unable to capture either purely rotational tasks (Motor Task 3) or fully combined translational and rotational movements, as seen in the final experiment (Motor Task 4). In particular, all models reviewed in this chapter entailing a purely rotational setting (Motor Task 3), were insufficient in both pointing and manipulation types of tasks. However, the most critical experiment was Motor Task 4 whereby both translational and rotational movements were combined, with varying spatial arrangements (directions and inclinations) under one setting, rendering the last experiment a full 3D task. In the final experiment, where the most complex spatial settings in 3D were effectively investigated, all compared models proved to be insufficient for both pointing and manipulation types of tasks. These observations from the results led the author to propose a novel metric to overcome these limitations. The derived model

Table 4.6: **The fitting results ($\mathbf{r^2}$) of all models across all motor assessment tasks.** Bold numbers represent the best model in the column group. For Motor Tasks 3 and 4, the existing models that are compared are in theory incompatible for combined movements of translational and rotation movements. Nevertheless, to allow for a fair comparison, as with the work of Stoelen and Akin (2010), these model extensions were adjusted in this work to accommodate for rotational motions. Consequently and as an example, Fitts' original formulation defined as $ID_t = log_2(\frac{2A}{W})$ was adjusted as $ID_r = log_2(\frac{2\alpha}{\omega})$. In this way, both IDs are added for Motor Task 4 and separate translation and rotation as two different indices of difficulty.

Fitting (r^2)	**Motor Task 1** (Translation Only)		**Motor Task 2** (Translation w/ Dir. & Inc.)		**Motor Task 3** (Rotation Only)		**Motor Task 4** (Trans. w/ Rot. & Dir. & Inc.)	
	Pointing	Manipulation	Pointing	Manipulation	Pointing	Manipulation	Pointing	Manipulation
Fitts	0.838	0.778	0.832	0.579	0.663	0.604	0.459	0.748
Hoffmann	0.862	0.869	0.789	0.511	0.659	0.600	0.569	0.762
Welford	0.852	0.776	0.832	0.577	0.673	0.610	0.465	0.755
MacKenzie	0.859	0.773	0.831	0.575	0.680	0.613	0.469	0.758
Murata and Iwase	0.859	0.773	**0.845**	**0.585**	0.680	0.613	0.476	0.759
Cha and Myung	0.862	0.869	0.833	0.574	0.659	0.600	0.578	0.770
Derived Model	**0.887**	**0.877**	0.788	0.509	**0.901**	**0.804**	**0.805**	**0.803**

outperformed other extensions in Motor Task 1 and 3 and most importantly in Motor Task 4, as shown in Figure 4.6 and Table 4.6.

From the results at hand and from the related work, it can be inferred that the derived metric could be used to assess human movement in teleoperation entailing 6DoF as commonly seen in robotics and VEs (O'Hara, 1987), especially with the use of MR technologies (Fani et al., 2018). From these results, it can be supported that the derived formulation, detailed in Section 4.5, can capture the complex spatial settings in 3D space associated with such scenarios as seen across all designed experiments (Motor Tasks 1,2,3 and 4). In addition, as the derived formulation is based on Fitts' law and by extent combines spatial and time-based metrics under a single formulation, further studies on human ergonomics, HCI and Human-Robot Interaction (HRI) could benefit in terms of standardisation as the result of increased inter-study comparability (Triantafyllidis and Li, 2021a; Steinfeld et al., 2006).

Table 4.7 summarises the main findings from this research work and outlines several

recommendations for task design entailing the quantification of upper-body human motor movements. These recommendations and guidelines are applicable for 3D user interfaces, VR-based interactions and robotic teleoperation.

4.6.2 Importance of Quantifying Combined Motor Movements

From the literature review in earlier parts of this thesis, it can be inferred that combining the two separate types of motion of translation and rotation is fundamental to cover most human movements as seen during interactions with virtual simulation environments and robotic teleoperation (Stoelen and Akin, 2010; Triantafyllidis and Li, 2021a,b; Triantafyllidis et al., 2020).

From the results of this work, it was observed that for fully combined movements, even by incorporating and adding the indices of difficulty IDs for both translation and rotation as with the model extensions of Stoelen and Akin (2010) and Kulik et al. (2020), the models are insufficient. The work of Kulik et al. (2020) observed a linear fitting of $r^2 = .780$ when combining both IDs, but limited in 2D movements across one line without directions or inclinations. However, it is worthwhile to point out that direct comparisons between different studies are rendered difficult. In particular, correlation coefficients (r^2) are highly influenced by the involved data points recorded and the different experimental settings involved (consult Figure 2.1). This, in turn, limits inter-study comparability (Triantafyllidis and Li, 2021a; Drewes, 2010). This limitation is explicitly why this work compared these models on the experimental data directly instead of comparing these across other experimental settings from the related work, as summarised in Table 4.6 and Figure 4.6 to mitigate to some extent the comparability limitations otherwise faced with.

Across all four designed motor task experiments, the most widely used model extensions of Fitts' law to date were compared. From the results it is shown that the proposed derived model demonstrated better fitting results than the existing extensions both in Motor Task 1 and 3, and most importantly when both translational and rotational movements were combined (Motor Task 4). In Motor Task 4, the derived formulation was better fitting for both pointing ($r^2 = .805$) and manipulation ($r^2 = .803$) types of task. All other existing formulations from the related work that were compared against the derived model, proved to be insufficient when attempting to assess and quantify both translational and rotational motions in full 3D space as commonly seen in real life.

The Final Derived Model and its Attributes

The superior fitting of the derived model compared to existing model extensions of Fitts' law assessing both translational and rotational motions in full 3D space can to a great extent be attributed to **three major factors.**

Firstly, it was observed from the results that equally weighing both IDs of rotation and translation is not supported. In particular, translational motions appear to be significantly easier than rotational motions as observed by the higher completion times of all involved participants. This implies that assuming that both translation and rotation have a linear effect on MT is not supported by the analysis stemming from this research study. Instead, from the results it can be inferred that it is crucial to separate these two terms as these should be treated as two separate types of motion. The separation of these two terms is accomplished in the derived model of this work by introducing different constants for each as explained in more detail in Section 4.5.

Secondly, the rotational tolerance ω also defined as the rotational accuracy, shared a non-linear relationship with MT in this study, in contrast to a linear relationship as with the work of Stoelen and Akin (2010). This difference can be attributed to a great extent to the higher degrees of difficulty required to match the rotation of an object associated with the 3D experiment in this work compared to 2D space as with Stoelen and Akin (2010). Hence the incorporation of ω in the model is of importance.

Thirdly, it can be concluded and inferred that from the 2D formulations, Hoffmann's formulation (Hoffmann, 1995) appears to present higher model fittings to the experimental data of this work, indicating the necessity of adding the size of the object F, hence why it is included in the final model. The importance of including the object size was presented in the manipulation types of tasks which took an overall longer time to complete except in purely rotational tasks (Motor Task 3), where pointing proved significantly more difficult, especially with the lowest target tolerance at $\omega = 2.5°$. Overall, this time discrepancy can partially be explained by the duration spent on grasping the object. It can hence be concluded that both W and F inversely affect MT and should therefore be included in a model as increased dimensions of those variables correspond to significantly decreased completion times.

The Negligible Effects of Directional and Inclination Angles

No notable, across either four motor tasks, evidence was found to support the consistent significance of either directional or inclinational angles. A significant effect of directional angles towards MT was only revealed in Motor Task 2 and only for pointing tasks ($p < .05$). Contrarily, inclination angles did present a significant influence towards MT with higher inclines corresponding to longer timings in Motor Task 2 ($p < .05$). However, this was later shown to have an insignificant effect on the final evaluation with combined movements in Motor Task 4. In the final combined motor task experiment, it was observed that translational and especially rotational variables influenced completion time MT significantly more than inclinations or directions. This implies that the difference between those two concepts of motion perhaps overshadowed any marginal influence that directional or inclinational angels could have had towards MT. Nonetheless, despite the negligible effect of translation and rotation on MT, this research still revealed marginally worse performance in front-to-back (90°, 270°) rather than lateral i.e. left-to-right (0°, 180°) movements, which is line with Barrera Machuca and Stuerzlinger (2019).

4.6.3 Limitations and Future Work

The limitations and future work of this study include the following:

- *Type of Movements* – Assessing human motor performance is complex due to the associated movements involved (Triantafyllidis and Li, 2021a). In particular, in this work, the influencing spatial variables and their effect on movement time were investigated across different motor task experiments with increasing complexity for upper-body movements. However, it would be beneficial to also extend this study to non-stationary manipulation tasks with the inclusion of navigation i.e. walking and lower-body movements. Moreover, the inclusion of movements stemming from the torso and/or shoulders would be worthwhile investigating as well.

- *Biomechanical Variations* – The individual biomechanical variations of participants are another important factor to be taken into consideration. An important variation identified for this particular work was the placement of objects and whether these were near the dominant hand of the participants. This was explicitly why this work retained consistency during recruitment and limited it to right-handed participants only. Further investigation on the handedness of individuals may prove important.

- *Multitude of Different Input Devices* – The plethora of different input devices, hand types, the variations in robotic end-effectors and the grasping types one can utilise may also affect the formulation of human models (Triantafyllidis and Li, 2021a).
- *The Complexity of Human Perception and Data* – As identified in the literature review in earlier parts of this thesis, human perception is a highly complex phenomenon and subject to each individual's exposure to the relevant technologies of a given experiment. Moreover, personality-related factors render the generalisation of human models for motor tasks more difficult (Triantafyllidis et al., 2020; Triantafyllidis and Li, 2021a). While mitigating this type of limitation is quite intricate, the recruitment criteria of participants can be clearly stated in related work to help alleviate some of these factors (Triantafyllidis and Li, 2021a,b).
- *The Scope and Limitation of Spatial Variables Investigated* – Due to the already exhaustive amount of variations of the spatial variables in 3D space introduced in this work, the investigation of those variables was limited to a maximum of four levels. It is advisable to investigate and study a wider range of spatial variables and extend this research, as it may perhaps shed additional light on how these affect task difficulty. For instance, replacing the cube object with a more complex 3D shape, such as a toy car, would allow for a wider range of rotational variations.
- *The Plethora of Choices for Defining Distances During Motor Movements* – This work derived a novel metric capable of measuring combined translational and rotational motor movements. Nonetheless, when satisfying spatial criteria concerned with combined movements, humans generally attempt to match rotational motions while simultaneously completing translational movements. For instance, in a pick and place task that necessitates to also match the orientation of an object, the satisfaction of the spatial orientation of said object is usually done while transporting it. In particular, the adjustment of the object's orientation is integrated into the movement, occurring predominantly towards the end of the transportation phase. The research conducted in this work was limited to a human study whereby matching the orientation of the object was either done before or after placing it on the target. Future investigations should delve into this nuanced aspect of combined movements and the measure of available distances one can employ. Consequently, future work needs to consider this important distinction during combined movements, the choices of distances and whether a new formulation is necessitated. Lastly, the

utilisation of quaternion over Euler angles in a formulation may further increase the generalisation of the derived model.

Additional expansion and concluding remarks of this chapter can be found in Chapter 7. In this work, to the best of the author's knowledge, the most widely used model extensions of Fitts' law were studied to derive a human performance metric by comprehensively analysing all spatial variables in full 3D space. Four increasingly complex motor task experiments were conducted to understand which spatial variables contributed to an observable influence on task complexity during motor tasks via a comprehensive statistical analysis. It was concluded that the derived metric can model combined translational and rotational motions in full 3D space (6DoFs), entailing the manipulation and pointing of objects, notably better than existing models. Moreover, as the model is based on Fitts' law, the time required to complete a set of motor tasks based on the spatial layout of the scene can be predicted. Essentially, the derived model combines both time and spatial-based metrics under a single formulation, increasing subsequent inter-study comparability. Consequently, these findings underline the value of the proposed model and its ability to quantify commonly observed intricate upper-body human motor movements.

Table 4.7: **The summary of the main findings of this work and in particular on the motor movement and quantification of motor task performance.** The table summarises the main results and implications for simulated teleoperation settings with recommendations surrounding the design of motor tasks and 3D interfaces.

Research Findings and Implications of Results
Fitts' original formulation can be extended towards 3D, however, does not adequately model or cover all of the associated spatial settings in higher-dimensional space. Hence, the formulation requires a clear definition of the distances associated with 3D.
Both translational and rotational types of motion are not only considered critical when attempting to quantify full 3D movements, yet should also be classified as two separate concepts and quantities. In the motor task experiments, it was observed that rotation predominantly influenced MT over translation. This influence was to an almost $\frac{2}{3}$ degree of the total contribution. Consequently, a new model was derived entailing a constant for each of those two motions, as these, as observed from the data, do not contribute an equal weight towards MT.
Pointing and manipulation types of interactions are perceived as inherently two very different types of interactions. Nonetheless, both of those types of interactions do follow and can be modelled by Fitts' original law, even in higher-dimensional 3D space.
The size of the manipulated object size matters. During pointing types of tasks and especially during physical manipulation of objects this holds true. For manipulation, the object size is an inseparable part of the metric. Adding the object size F into the equation as suggested by Hoffmann's extension (see Hoffmann (1995)) of Fitts' law is feasible and confirmed in the derived model extension of this work.
The rotational tolerance ω, or simply the rotational accuracy, predominately affects MT over all other spatial variables when combined movements are initiated. This was particularly observed in cases of very low rotational tolerances (2.5°). Consequently, strict rotational accuracy requirements during object placement (i.e. teleoperation), which by extent correspond to low rotational tolerance values, are very likely to correlate to a notable increase in operator completion timings.
Motor-Task Design Recommendations in VR Interfaces and Teleoperation
"Strict" accuracy requirements should be avoided as low rotational tolerances, i.e. high rotational accuracy, contribute to high MT, following an evident non-linear relationship from the data of this work.
Motor tasks that require moving objects towards or away from the view direction of the user should be avoided if possible. From the results of this work, it was observed, even to a marginal degree, that front-to-backwards movements take notably longer time to complete than left-to-right ones and vice versa.

Chapter 5

The Robotic Manipulation Network – Biologically-Inspired Intelligence

To this point, two important distinct goals were presented, namely how different combinations of multisensory feedback affect the human sensory-motor system in generating meaningful motor movements and actions (Chapter 3) and in what ways these upper-body biomechanical movements can be quantified (Chapter 4). From the aforementioned advances in sensory-motor perception, it was inferred that humans are incredibly dexterous and proficient in interacting with their surrounding environment.

In a broader context and from a biological standpoint, by virtue of the incredible sensory, dexterous and cognitive abilities humans possess, adapting to environmental demands appears to be anything but a difficult concept (Ashe et al., 2006; Ortenzi et al., 2019). Focusing on a narrower scope, by virtue of these abilities, humans adapt to their environment entailing complex sequences of motor tasks in a seamless manner (Triantafyllidis et al., 2020; Billard and Kragic, 2019). However, when switching to an embodied intelligence perspective such as robots, achieving the same level of physical interactions as humans in long-horizon tasks entailing a plethora of intricate sequences is currently far from trivial (Tee et al., 2022; Billard and Kragic, 2019; Davchev et al., 2022). Ultimately,

Parts of this chapter are published as: E. Triantafyllidis, F. Acero, Z. Liu and Z. Li. (2023). "Hybrid hierarchical learning for solving complex sequential tasks using the robotic manipulation network ROMAN". In: *Nature Machine Intelligence*, Volume: 5, Issue: 9, Pages: 991–1005, DOI: 10.1038/s42256-023-00709-2.

comprehending the biological processes that govern these abilities and render humans so adept at engaging with their environment, might pave the way for emulating such human-like thinking and skill in machines. Consequently, it is natural to imagine how machine intelligence and more broadly machine learning approaches could benefit from drawing inspiration from a biological perspective.

As part of the penultimate contribution of this thesis and inspired from a biological standpoint from Chapter 3 and Chapter 4, this chapter presents a biologically-inspired learning architecture – the **RObotic MAnipulation Network (ROMAN)**. ROMAN is a Hybrid Hierarchical Learning (HHL) event-based architecture composed of an ensemble of various specialising in manipulation skill experts, centrally governed by a supervising gating network referred to as the Manipulation Network (MN). ROMAN is designed to address notably intricate long-horizon sequential robotic manipulation tasks. By combining the exploratory nature of RL and simultaneously exploiting the higher-level cognitive skills of humans in the form of imitation and supervised learning (via GAIL and BC), ROMAN achieves task versatility and adaption in intricate robotic manipulation settings. This hybrid combination of existing learning paradigms alongside the higher-level task decomposition of experts within the hierarchy renders ROMAN particularly useful in intricate robotic manipulation tasks. Results show that ROMAN demonstrates robustness in tasks necessitating the timely and correct sequential coordination of a plethora of distinct specialising experts in order to solve long-horizon end-goals. Validated against (i) monolithic equivalents sharing an identical hybrid training procedure, (ii) extensive ablation of the internal learning paradigms and (iii) evaluated against increasing levels of uncertainty and horizons, ROMAN continuously and consistently maintains robustness. Most importantly, by virtue of the employed HHL, ROMAN exhibits beyond demonstrated behaviour by its dynamic ability to meet environmental demands featuring the architecture's autonomous failure recovery capabilities and adaptation to local minima.

Collectively, the results of this chapter underline the value and potential of the biologically inspired hierarchical task learning ROMAN architecture for intricate, long-horizon manipulation tasks necessitating the correct orchestration and activation of a plethora of adaptive motor skills. Figure 5.1 visually depicts the ROMAN architecture in addition to its distinct specialising experts. This chapter will elaborate on the overall derivation, technical details and results of ROMAN, while Chapter 7 presents the limitations and future work in more detail.

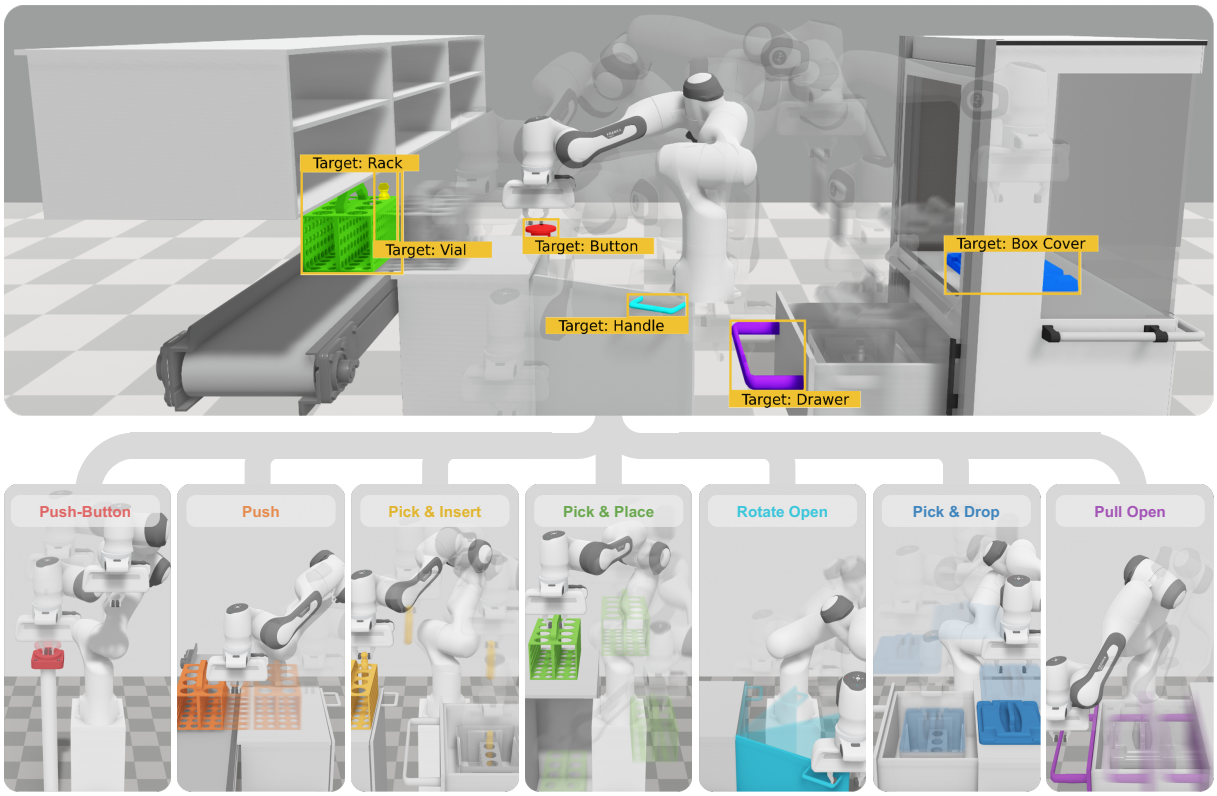

Figure 5.1: **The capabilities of ROMAN:** A Hybrid Hierarchical Learning (HHL) framework for hierarchical task learning, capable of solving long-horizon tasks that necessitate the successful sequential activation and coordination of diverse expert skills to solve a sequence of non-interrelated tasks, commonly encountered in robotics and physics-based interactions. The derivation of high-level specialised experts in ROMAN, allowed the conceptualisation and derivation of a gating network, referred to as the Manipulation Network (MN). The MN is supervising a plethora of expert NNn, treated as macro-actions, trained for elevated task-level scene understandings, the planning of complex sequential long-time horizon tasks and the successful and timely activation of the incorporated specialising expert networks in the hierarchical architecture. A set of seven specialised manipulation skills were derived, that are common in daily life (Billard and Kragic, 2019; Kumar and Todorov, 2015), which can be combined to create and solve a notably intricate sequence of manipulation skills. The specialised skills in ROMAN include (i) *Pushing a Button*, (ii) *Pushing*, (iii) *Picking & Inserting*, (iv) *Picking & Placing*, (v) *Rotating-Opening*, (vi) *Picking & Dropping* and (vii) *Pulling-Opening*. Contrary to conventional planning methods or state machines, ROMAN exhibits adaptability in (i) randomised task sequences, (ii) generalisation beyond demonstrated behaviour, as well as (iii) recovery and robustness against local minima. The capability of the framework to achieve such versatility is attributed to (i) the employed HHL in ROMAN's core architecture and (ii) the high-level task decomposition of complex sequences by the various experts in the architecture, allowing the central MN to be trained on high-level scene understanding and the orchestrations of experts. The system architecture is based on a MoE, capable of successfully adapting to environmental demands, overcoming various levels of uncertainties and most importantly learning with minimal human imitation.

Chapter Outline – The chapter is structured in four main sections. First, Section 5.1 presents the motivation for using human-inspired skills in the form of imitation, to align embodied intelligence closer to the capabilities exhibited by humans from a biological standpoint. Moreover, this section outlines the importance of hierarchical architectures as well as frameworks and ultimately how to address notably intricate, long-horizon sequential tasks in robotics and manipulation, which represents itself as a gap in existing work. In Section 5.2 the methodology is introduced, presenting the system apparatus, the derivation and full characteristics, including the architecture, of the ROMAN architecture as well as the learning paradigms therein. Section 5.3 presents the results of the ROMAN architecture, tested and evaluated against different spatial complexities, environmental uncertainties and existing state-of-the-art learning approaches and methods. Finally, in Section 5.4, the results from the ROMAN architecture are discussed and framed with existing approaches. The discussion entails how the HHL approach rendered the whole architecture robust and capable of dealing with uncertainty and cases beyond those encountered in the demonstrated sequence and ultimately summarises, in brief, future work and improvements. Chapter 7 discusses in more elaborate detail the limitations and future work for ROMAN and in a more broader context dealing with intricate and long horizon tasks.

5.1 Introduction

From a biological perspective, humans interact with their surrounding environment with seemingly minimal effort, even when performing notably intricate in-sequence tasks (Triantafyllidis and Li, 2021a; Ashe et al., 2006; Ortenzi et al., 2019). Even though some of these tasks are so diverse in their nature, by virtue of their highly complex cognition, perception and motor dexterity, solving intricate sequences of tasks appears to require very little effort for humans (Triantafyllidis et al., 2020; Billard and Kragic, 2019).

From an embodied intelligence perspective, solving a sequence of multiple complex tasks over a long time horizon is notably challenging, especially in robotic manipulation (Tee et al., 2022; Billard and Kragic, 2019; Davchev et al., 2022; Fox et al., 2019). Most notably, a task as simple as retrieving a glass from a shelf, pouring in a liquid and placing it onto a table may seem a trivial task, but from an embodied intelligence perspective, remains significantly challenging (Triantafyllidis et al., 2023b). In particular, successful sequential manipulation is achieved when (i) high-level skills are satisfied, (ii) sensory

events are predicted, (iii) the end goals are known and (iv) the sequences of different skills are conceptualised in the minds and more broadly by the nervous systems of humans (Flanagan et al., 2006; Ortenzi et al., 2019).

Despite these limitations, robots as embodied intelligence can perform repetitive manipulation tasks with high precision, rendering these notably valuable for certain settings, provided these are confined to specific tasks (Triantafyllidis et al., 2020; Zhang et al., 2021). Some of these tasks include picking and placing (Zhang et al., 2018b; Triantafyllidis et al., 2020), swing-peg-in-hole (Chebotar et al., 2019; Lee et al., 2019), catching in-flight objects (Schill et al., 2015), insertion (Lee et al., 2019; Schoettler et al., 2020) or solving something as complex as a Rubik's cube (Andrychowicz et al., 2020). However, when it comes to solving a sequence of multiple intricate tasks, notable challenges arise (Zhang et al., 2021).

Consequently, given the advanced cognitive behaviour and motor skills possessed by humans, it is natural to seek effective techniques to align these skills closer to AI and derive embodied intelligence inspired from a biological perspective to address notably intricate scene understandings for complex robotic manipulation tasks. Solving complex, long-horizon sequential tasks during robotic manipulation remains a challenging area of research (Davchev et al., 2022; Fox et al., 2019). While current deep RL algorithms have shown promise in learning manipulation-based skills, they require millions of state transitions and may not converge with stability, particularly in complex and sparse reward tasks (Fox et al., 2019; Marzari et al., 2021). Machine intelligence is presently limited in providing critical decision-making skills and inferring scene understandings, which are unique and highly complex cognitive abilities offered by humans (Goldberg, 2019; Triantafyllidis et al., 2020). In this chapter, a novel approach of harnessing human imitation of highly specialised skills in a student-teacher fashion is presented – ROMAN, which is tailored to address notably intricate, long-horizon sequential manipulation tasks, commonly encountered in the everyday lives of humans and routinely seen in robotics (Billard and Kragic, 2019; Triantafyllidis and Li, 2021a; Triantafyllidis et al., 2021; Triantafyllidis et al., 2020).

5.2 Methodology

In this section, the internal architecture of ROMAN is described and all its components are analysed, with a focus on the derivation of the architecture. The main characteristic of the ROMAN architecture is the employed HHL approach and the high-level task decomposition of complex sequences. Within its architecture, ROMAN is composed of a plethora of different expert skills each specialising in diverse and essential types of manipulation tasks. When these experts, in turn, are activated in their correct sequence by a supervising gating NN, known as the MN, complex long-horizon sequential tasks can be solved. ROMAN was evaluated in a simulation environment consisting of realistic physics-based interactions to allow a close approximation of the simulated architecture to realistic use cases entailing different types of manipulation tasks that are commonly seen in robotics (Billard and Kragic, 2019).

It is worthwhile to point out that the main contribution of ROMAN is not necessarily the capabilities exhibited by the derived experts, but rather how this hybrid hierarchical architecture can be used to compose the experts and effectively solve intricate sequential tasks in 3D space, characterised by notably long time horizons and non-interrelated sequential sub-sequences. The main contribution of ROMAN is its ability to orchestrate in a high-level manner, an arbitrary amount of specialising expert skills to solve intricate long-horizon sequential manipulation tasks with varying amounts of sensory uncertainty while exhibiting intelligent and adaptive behaviour.

5.2.1 System Overview

A significant part of the system apparatus stemmed from the previous work presented in earlier chapters of this thesis. The derived hierarchical architecture is validated upon an intricate simulated medical laboratory setting entailing (i) careful handling of small objects, (ii) the necessity to perform multiple tasks and (iii) the correct sequence of tasks to complete a long and complex end goal. The design and derivation of the environment were accomplished in such a way as to derive as many compositional sub-tasks as possible. In this environment, the value of a hierarchical architecture necessitating a plethora of specialising experts and the employment of an appropriate task decomposition can be highlighted in a significantly intricate long-horizon sequential setting.

To achieve the aforementioned, the simulation engine Unity3D was used, containing the

built-in NVIDIA physics engine (PhysX 4.1). To allow the training of each component of the architecture of ROMAN, the simulation was coupled with the PyTorch-based ML-Agents toolkit, to interface the simulation with learning-based approaches [1], consult Juliani et al. (2018). To ensure stable and reliable physical modelling of the robotic system, the ROS# plugin was incorporated to import physics models of robots and objects (URDF) into the Unity3D engine. The physics simulation time-step, or frequency, was set at 1,000Hz to ensure robust and stable physics performance, including realistic physical properties and frictions (Triantafyllidis et al., 2021; Yang et al., 2020). It was deemed critical to ensure stable and realistic physics performance as combining deep RL methods with realistic simulators is advantageous for training intelligent agents as shown by the state-of-the-art and the promising capabilities of Sim2Real transfer (Lee et al., 2019; Yang et al., 2020). The rationale of using a simulator was deemed particularly important due to its promising capabilities of performing a large number of iterations, allowing for a notably easier and quicker interaction of the agent and its environment, while also minimising the risk of physical and hardware damage when concerned with robotics (Triantafyllidis et al., 2021; Lee et al., 2019; Yang et al., 2020).

Furthermore, a low-pass filter was implemented for filtering the input signals, i.e. exteroceptive state vector to the NNs, The control loop of each expert NN was set to 25Hz, with the exception of the gating NN (the MN) which operated at 10Hz. The robotic system used for the experiments was the Franka Panda Emika robot, with its default gripper configuration. All simulated tasks were within the robot's workspace. All expert NN control policies, from this point onward presented, operated in the direct velocity control of the end-effector gripper. An Inverse Kinematic (IK) solver was used to solve the kinematic chain to the end-effector position. The system overview entailing the main simulation environment and the overall depiction of the ROMAN framework are shown in Figure 5.2. The parallelism in the training procedure of the robot allowed for concurrent training instances, which in turn significantly boosted training times. The training times varied between the experts and the gating network significantly. For more details regarding the complete timings consult Table 5.3, with additional details presented further along this section. For more expansion on the system details, consult the Appendices, Section A.1.

[1] https://github.com/Unity-Technologies/ml-agents

5.2.2 Internal Architecture and Learning Paradigms

The internal architecture of ROMAN utilises two IL algorithms, Generative Adversarial Imitation Learning (GAIL) (Ho and Ermon, 2016), as well as Behavioural Cloning (BC) (Torabi et al., 2018). These two imitation / supervised learning paradigms are then coupled with the Reinforcement Learning (RL) algorithm Proximal Policy Optimization (PPO) (Schulman et al., 2017). The fusion of these paradigms was inspired from a neuro-biological standpoint, which ultimately allowed the architecture to successfully and robustly approximate, closer to the capabilities exhibited by humans in the form of imitation skills and the exploratory nature of RL, intricate long-horizon sequential daily activity tasks. This fusion alongside ROMAN's general architecture is tailored for autonomous robotic operation necessitating adaptive motor skills.

Behavioural Cloning (Warm-Starting the Policy)

As a first instance, BC was utilised until a given number of initial epochs, to help warm-start the policy of each involved network within ROMAN. While BC is a compelling approach, the use of exclusively utilising supervised learning throughout the training process was avoided to allow the agent to explore further samples. The reasoning behind this decision was to improve upon what was demonstrated and retain a small demonstration dataset (Ho and Ermon, 2016; Zhang et al., 2018b). This is confirmed by the literature review highlighting that the sole dependence on using BC limits overall generalisation to out-of-distribution states, effectively limiting it almost exclusively to the trajectories found in the demonstrations provided (Ho and Ermon, 2016; Reddy et al., 2020). Moreover, even from a biological standpoint "naive" imitation is misguided, as even humans in their early lives rely on learning useful representations from unlabeled data entailing trial and error, as suggested by Zaadnoordijk et al. (2022). Switching back to a learning perspective, sole reliance on BC can negatively affect the policy and can lead to drifting errors when the agent encounters new trajectories that are outside of what was demonstrated upon (Ho and Ermon, 2016; Codevilla et al., 2019). Consequently, in line with previous work concerned with neuro-biology and embodied intelligence, sole dependence on BC should hence be avoided. Inspired by the aforementioned, in this work, a human demonstrator imitates a dataset entailing state and action transitions s_t^d, a_t^d and by implementing BC within the internal architecture of the ROMAN framework, a NN policy $\pi(s_t) = a_t$ using supervised learning is trained to minimise the Mean Squared Error (MSE)

loss between a_t^d and a_t for the demonstration dataset.

Generative Adversarial Imitation Learning (Exploitation)

Unlike the limitations of using exclusively BC and to effectively match the provided human demonstration data over a period, also known as a horizon, inverse RL and in this case GAIL was used (Ho and Ermon, 2016). GAIL was used after the warm-starting of the policy by BC's cutoff point and thereafter GAIL commenced and was active throughout training. The reasoning for using GAIL was to attempt to minimise and approximate the divergence between the agent's policy and that of the demonstrator making use of exploitation. Practically, to allow the use of GAIL and update the corresponding policy parameters, a proxy imitation reward signal obtained by GAIL was instead used, which is utilised in the form of an intrinsic reward, described in more detail at the end of the current section.

With GAIL the divergence of the agent's and the expert human's policy can be minimised by sampling a set of agent (τ_A) and expert (τ_E) trajectories of the state and actions (s_t, a_t). The trajectories stemming from the human expert are sampled from a given demonstration dataset whereas the agent trajectories are sampled from a generative model also known as Generator (G). However, the generator instead of being rewarded solely by the environment via extrinsic rewards, is instead being rewarded by a scalar score provided by the Discriminator (D). This is achieved by implementing the discriminator as a separate NN. Within this process, the discriminator attempts to differentiate whether the trajectories are produced by the expert or the agent and in turn rewards the generator if that divergence between these trajectories is minimised. The discriminator can be thought of as becoming increasingly "stricter" over time, resulting in the generator, e.g. agent, improving its overall performance at imitating and converging towards the behaviour that was demonstrated by the human expert. This process can be formulated mathematically as follows:

$$E_{\tau_E}[\nabla log(D(s_t, a_t))] + E_{\tau_A}[\nabla log(1 - D(s_t, a_t))] \tag{5.1}$$

whereby the terms E_{τ_E} and E_{τ_A} represent the expert and agent trajectories from the training respectively. Both of these terms are represented as inputs to the discriminator network (D). The discriminator, in turn, outputs a continuous value ranging from 0 to 1, with a value closer to 1 corresponding to the agent, or generator, resembling and approximating a trajectory closer to that of the expert's. Consequently, the higher this value

becomes, the less the divergence between the two trajectories and by extent corresponding to higher amounts of imitation. Hence, the discriminator can be used as a reward signal to train the generator with the goal of the latter to mimic the expert's demonstrated data.

From the above, it can be observed that with GAIL, both states and actions of the agent and the expert need to be minimised as to maximise imitation. However, in the case of the ROMAN framework, GAIL is modified to only allow the discriminator to distinguish between the states (s_t) but not the actions (a_t) of the demonstrated trajectories. The rationale behind this decision is to encourage the policy to generate its actions, instead of attempting to mimic those as well, with the goal of increasing exploration instead of purely mimicking the expert. As it will be seen further along this chapter, this slight modification to GAIL, when coupled with RL via extrinsic rewards, will encourage ROMAN to adapt to beyond demonstrated behaviour and recover from local minima when these are encountered. Consequently, Equation 5.1 is reformulated as:

$$E_{\tau_E}[\nabla log(D(s_t))] + E_{\tau_A}[\nabla log(1 - D(s_t))] \tag{5.2}$$

As such, sampling only the states via GAIL allowed the framework to be overall less restrictive in terms of imitation. As aforementioned, discriminating against both states and actions between the demonstrator and the expert trajectories as with the original formulation of GAIL (Ho and Ermon, 2016), would have potentially led to disallowing the agent to further explore other actions. These actions in turn could in actuality lead to better adaptation based on the state space instead of "naively" copying of identical imitation.

In addition to balancing exploration and exploitation via this approach and limiting BC to purely warm-starting the policy and solely depending on supervised learning, translated into a notably reduced necessary dataset. This dataset was notably smaller in size compared to the related work in the context of training agents over long and intricate sequential tasks (Zhang et al., 2018b; Marzari et al., 2021). In this way, the need for a large demonstration dataset stemming from a human can be mitigated, which is preferable as larger data introduces higher variability and can eventually lead to operator fatigue resulting in degraded performance over time (Triantafyllidis et al., 2020; Triantafyllidis et al., 2021). More importantly, in real-world applications, an overly end-goal-based learning may not comply with the standards of industrial procedures. This approach, from a real-world industrial perspective, takes a minimal set of demonstrations compliant with industrial standards and enables robots to extrapolate and acquire resilient skills based

on these demonstrations. This particular research problem matches well with the real-world settings, which renders IL and hierarchical learning approaches more attractive to be deployed in practice (Triantafyllidis et al., 2020, 2023b).

Reinforcement Learning (Exploration Beyond Imitation)

In addition to the aforementioned two IL approaches, task-related environmental rewards, referred to as extrinsic reward signals, as per the original formalism of RL are utilised. As aforementioned, the framework's policy is warm-started via BC up until a given epoch and thereafter intrinsic rewards are utilised via GAIL. To allow the agent to explore further than what was demonstrated, the use of extrinsic rewards was incorporated to provide a small contribution towards the final policy to avoid and overcome some of the limitations of exclusively depending on pure imitation (Ho and Ermon, 2016; Ross et al., 2011).

While both intrinsic from IL as well as extrinsic task-related rewards to update the policy are utilised, the IL reward being scaled within ROMAN has the highest scaling factor and by extent being the main learning signal provider. This balance between exploitation, via imitation and intrinsic rewards, as well as exploration, via extrinsic rewards in a generic RL fashion, rendered ROMAN to positively adapt in cases not encountered during demonstrations and under the presence of uncertainty. In particular, further along, this chapter, Section 5.3 details how this balance between exploitation and exploration led to the capability of ROMAN to recover from local minima. This adaptation was showcased even amongst the most intricate sequence activations of experts that were either attributed to incorrect sequential activation or the result of errors stemming from an individual expert. The RL algorithm of choice for ROMAN is PPO due to its robustness and flexibility to various hyperparameter settings (Schulman et al., 2017). More details regarding the rationale of utilising PPO can be found in Section 2.3. Denoting the policy π_θ as a NN parameterised by weights θ, the PPO update at step k is given by

$$\theta_{k+1} = \arg\max_{\theta} \mathbb{E}_{s,a\sim\pi_{\theta_k}} \left[L(s, a, \theta_k, \theta) \right] \tag{5.3}$$

with a clipped loss function $L(s, a, \theta_k, \theta)$ that has a surrogate term, a value term and an entropy term, consult (Schulman et al., 2017).

Hybrid-Hierarchical Learning: The Fusion of BC, GAIL and RL

The ROMAN framework is capable of solving long and complex sequential manipulation tasks by "fusing" a set of learning algorithms for an effective balance of exploitation and

exploration. Foremost, supervised learning on the policy is performed using a demonstration dataset via the use of BC. In this process, the policy updates are driven by the corresponding MSE loss on the provided dataset. BC is used up until a given epoch to warm-start the policy and thereafter GAIL is used in the form of an intrinsic reward signal provided to the main PPO algorithm performing policy updates. Simultaneously, extrinsic rewards are utilised stemming from the agent's exploration and interaction with the environment following the default RL paradigm. Both intrinsic (via GAIL) and extrinsic (RL) rewards are provided to PPO acting as the general-purpose algorithm to perform policy updates. This, in turn, allowed the use of different reward terms for intrinsic r^i and extrinsic r^e. In particular, r^i are provided by the discriminator score from GAIL while r^e are provided by the environment as per the RL formalism.

The intrinsic reward term is defined as $r^i = -log(1-D(s_t))$, whereby $D(s_t) \in (0,1)$ denotes the score provided by the discriminator, which, in turn, acts as a proxy reward term that can be used by PPO to maximise the GAIL objective. When training simultaneously with both GAIL and RL, a linear combination of both reward terms is used such that $r = r^i \cdot \lambda^i + r^e \cdot \lambda^e$. The terms λ^i and λ^e represent fixed scaling parameters for extrinsic and intrinsic rewards respectively. With ROMAN's HHL control policy and architecture, more focus on the imitation i.e. on the intrinsic reward (r^i) was given rather than on the extrinsic reward (r^e). Consequently, the rewards for r^i are several magnitudes larger than for r^e and can be formulated as $\lambda^i > \lambda^e$. Using the latter reward combination, the returns are computed as the discounted sum of rewards and used to perform the PPO update on the policy as with Equation 5.3. This allowed ROMAN to get closer to the capabilities exhibited by humans and to be more closely aligned to a neuro-biological perspective as suggested by Zaadnoordijk et al. (2022). As such, the fusion of the aforementioned learning paradigms allowed the architecture to benefit from the priory knowledge humans possess during motor tasks (Goldberg, 2019), in the form of imitation, yet, also allowed the policy to further explore upon the imitated behaviour and potentially lead to additional improvements by virtue of the exploratory potential of RL (r^e) and GAIL's provided scalar score (r^i) with its modification shown in Equation 5.2.

The algorithmic depiction of the employed hybrid learning procedure used for all incorporated NNs (experts and MN) within the hierarchical architecture of ROMAN, is shown in Algorithm 1. Figure 5.2.a visually depicts the individual NN architecture of each expert while Figure 5.2.b depicts the higher-level architecture of the ROMAN framework and in

particular the orchestration of weights by the main MN. It is noteworthy to specify that in Figure 5.2.b, the exteroceptive information provided to each NN from the environment is dependent on the individual designated objective of each expert. Consequently, only the exteroceptive information relevant to the higher-level goal of that NN is provided. In contrast, the MN observes the entirety of the environment to allow for higher-level scene understandings and ultimately the orchestration of the plethora of entailed NN components treated as macro-actions.

Algorithm 1 Algorithmic Depiction of the Hybrid-Learning Procedure within ROMAN

Input: A policy network with parameters θ, a discriminator network with parameters θ_d

Behavioural Cloning to Warm-Start the Policy
for *epoch* in $[1, num\_bc\_epochs]$ **do**
    Collect expert trajectories
    trajectories $\leftarrow$ collect_expert_trajectories(env, expert_policy)
    Compute BC loss and update policy network
    loss_bc $\leftarrow$ compute_bc_loss(policy, trajectories)
    $\nabla_\theta \leftarrow$ compute_gradient(loss_bc, θ)
    $\theta \leftarrow$ update_weights(θ, ∇_θ)
end for

Reinforcement Learning (PPO) with Extrinsic and Intrinsic (GAIL) Rewards:
for *epoch* in $[1, num\_ppo\_epochs]$ **do**
    Collect trajectories using PPO with both r^e and r^i
    trajectories $\leftarrow$ collect_ppo_trajectories(env, policy, λ^e, λ^i)
    Train discriminator with both expert and generated trajectories, using states only (s_t)
    loss_discriminator $\leftarrow$ compute_gail_loss(discriminator, expert_trajectory, generated_trajectory)
    $\nabla_{\theta_d} \leftarrow$ compute_gradient(loss_discriminator, θ_d)
    $\theta_d \leftarrow$ update_weights($\theta_d, \nabla_{\theta_d}$)
    Update policy network with PPO using extrinsic and intrinsic (GAIL discriminator) reward
    loss_policy $\leftarrow$ compute_ppo_loss(policy, trajectories, discriminator, λ^e, λ^i)
    $\nabla_\theta \leftarrow$ compute_gradient(loss_policy, θ)
    $\theta \leftarrow$ update_weights(θ, ∇_θ)
end for

5.2.3 Vision System

As part of a preliminary investigation and to approximate a Sim2Real approach with the ROMAN framework, a vision system using an Red Green and Blue (RGB) camera

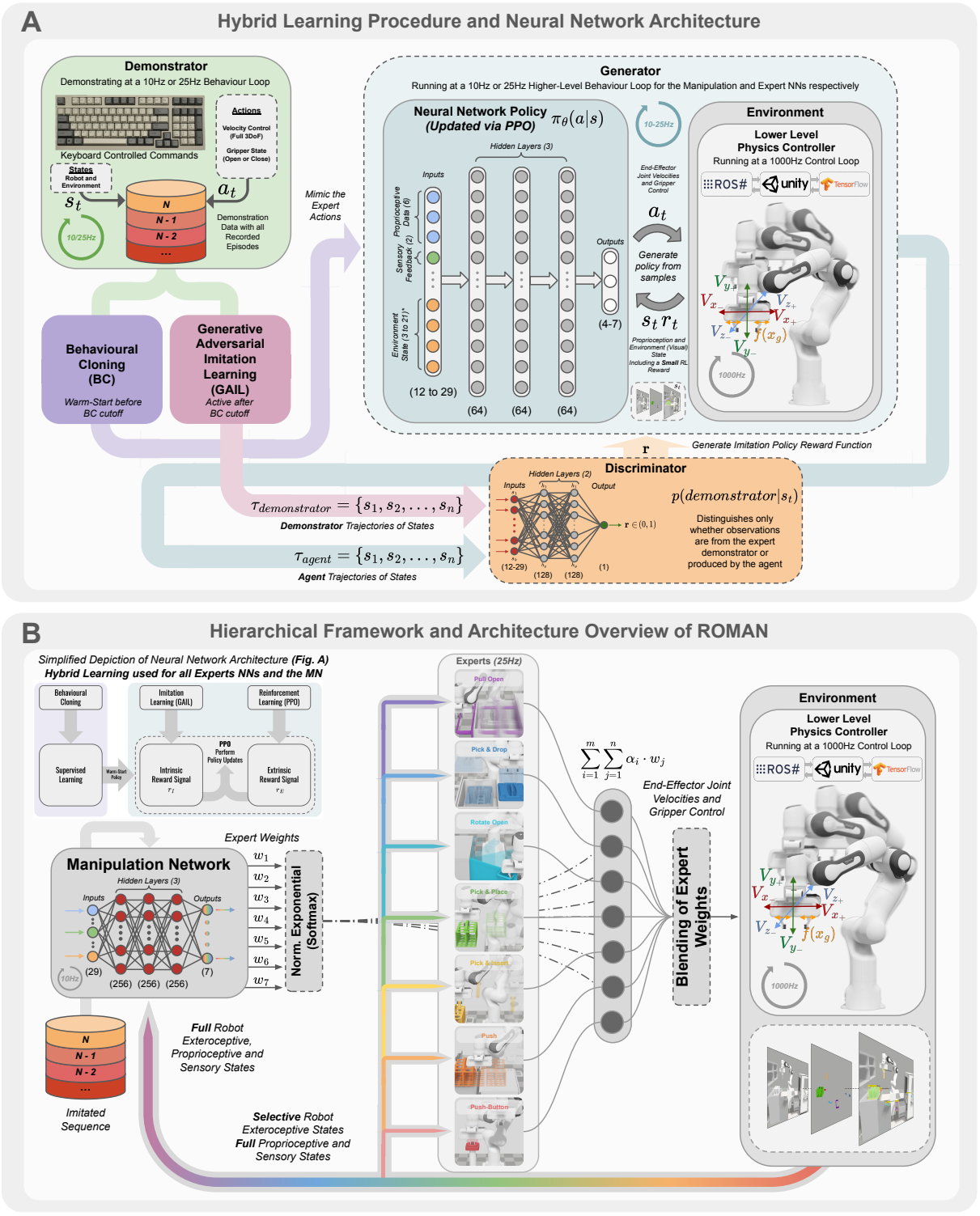

Figure 5.2: **The hybrid hierarchical architecture of ROMAN. Figure (A):** The employed hybrid learning for each NN within the hierarchical formation of the ROMAN architecture, including the incorporated high-level expert NNs and gating network (the MN). **Figure (B):** The higher hierarchical formation of ROMAN and how the experts are orchestrated and activated by the supervising MN. The Multi-Layer Perceptron (MLP) for the NNs are visually depicted in both figures.

was implemented. The camera served as an alternative and more realistic approach to providing exteroceptive information from the simulation by predicting the poses of the different Object of Interest (OI)s, as one would expect from a realistic robotic environment. The employed vision system implements an object detection and pose estimation module based on the VGG-16 backbone architecture, consult the works of Tobin et al. (2017). The system was initialised with pre-trained weights on the ImageNet dataset and fine-tuned using a custom dataset. The custom dataset was created by capturing the OIs from the simulated environment, including both the segmentation and labelling of the OIs. The output of the network predicted the poses of all OIs, in particular their 3D positions. During visual occlusions, in particular, the last known positions of the OIs were provided to the MN.

The rationale behind the decision to use a camera setup was to validate ROMAN's robustness in a realistic setting, whereby pose prediction errors are a common occurrence and visual occlusions naturally materialise. Nevertheless, since the pre-trained object detection module from the vision system attained variable levels of positional error (Tobin et al., 2017), additional increasing levels of Gaussian distributed noise were simulated to all exteroceptive observations of all NNs, in line with related work (Rajeswaran et al., 2018; Fox et al., 2019). This, in turn, allowed for a more comprehensive and quantifiable overview of ROMAN's robustness against increasing levels of Gaussian-distributed uncertainty. Ultimately, both the addition of exteroceptive uncertainties stemming from the vision system and the simulated Gaussian noise allowed the evaluation of ROMAN's resilience. Ultimately, as it will be shown further along this chapter, these underlined the necessity of a hybrid learning approach within a hierarchical architecture solving complex and long-horizon sequential tasks.

5.2.4 Human Demonstration Acquisition

The demonstration dataset was provided via a set of keybindings from a generic computer keyboard. In particular, the keyboard was utilised to provide **two levels** of demonstrations. **First**, the demonstrations were provided to the specialising expert NNs, with keybindings corresponding to the direct velocity control of the robotic end-effector as well as the binary state of opening or closing the gripper. The incorporated expert NNs in the hierarchical architecture of ROMAN shared identical actions and specialised in different manipulation skills. **Second**, given the pre-trained expert NNs, a demonstrated

sequence for the MN was provided via a set of different keybindings. The keybindings for the MN corresponded to the weight assignment of the incorporated experts in the hierarchical architecture. Hence, the expert demonstrations were specific to each individual expert's specialised skill and goal, which allowed these pre-trained networks to be thereafter coordinated by the MN's demonstrated sequence for the sequential activation of the task.

All demonstrations acquired were used to warm-start the policy via BC, as well as for GAIL which was employed in the form of an intrinsic reward alongside the PPO algorithm. In particular, for GAIL, the discriminator was tasked to differentiate whether the trajectories of those demonstrations stemmed from the generator (i.e. agent) or the human expert, providing, in turn, a scalar score ($(\in 0, 1)$), representing the intrinsic reward r^i. A total of $N = 20$ demonstrations were provided to pre-train the ROMAN expert NNs, and a total of $N = 42$ demonstrations were provided to the MN, corresponding to $N = 6$ for each of the seven sequential scenarios. The employed technical approach of using a keyboard for generating demonstration data and imitating trajectories for the expert NNs and the MN, as well as the corresponding technical implementations, to the best of the author's knowledge, did not stem from related published works.

In terms of visual feedback for the human expert, a generic 2D display monitor was provided to visually observe the task and imitate the desired behaviour. Moreover, two cameras in an orthographic projection were rendered onto the monitor to allow the depiction of the simulation environment from an upper and side-view perspective. Due to the lack of a stereoscopic display and perspective, the addition of two different camera perspectives allowed, to some extent, to overcome the issues that arise when determining depth-associated distances in the simulation, in line with previous work (Triantafyllidis et al., 2020; Barrera Machuca and Stuerzlinger, 2019; Triantafyllidis and Li, 2021a). A promising alternative would have been to directly project the environment visually on either 3D displays or MR technologies similar to related work on robotics (Triantafyllidis et al., 2020; Zhang et al., 2018b). The use of stereo displays as the means of imitating behaviours from human experts to intelligent agents remains part of the future work of this research. The derived sub-tasks that form each sequential task based on the medical/laboratory setting which ROMAN was evaluated upon are detailed in Table 5.1.

Table 5.1: **The individual decomposed tasks that are derived from the environment, with their sub-task goals described and their interdependencies listed.** A successful sequential task is defined as successfully having met all seven of the outlined sub-task goals. A single failure amongst the sequence of the below-mentioned tasks is treated as an overall error in the analysis of the results.

	Manipulation Task and Goal	Dependency	Scenario Case
(i)	Open drawer to expose container encasing vial	-	Scenario 7
(ii)	Unbox container to expose vial	(i)	Scenario 6
(iii)	Open cabinet to expose rack	-	Scenario 5
(iv)	Place rack on top of cabinet	(iii)	Scenario 4
(v)	Retrieve vial and place it onto rack	All of the above	Scenario 3
(vi)	Push rack containing vial onto conveyor	All of the above	Scenario 2
(vii)	Activate conveyor by pressing button	All of the above	Scenario 1

5.2.5 Characteristics of the Task

The built-in physics engine PhysX within Unity3D allowed the derivation of numerous and diverse manipulation tasks all containing physical properties and advanced physical characteristics. Some of those physical characteristics included hinges, linearly moving objects and spring joints. Complex mechanical tasks such as unlocking a lock for example were disregarded due to their notably difficult design as well as the lack of relevance for the overall validation of the system. The full sequence of manipulation tasks as seen in the simulation environment is visually depicted in Figure 5.1. The full sequence decomposed into its relevant sub-tasks is shown in Figure 5.5.d.

The task was conceived and inspired by a medical laboratory setting. In this setting, frequently encountered manipulation tasks are present whereby these often involve a plethora of different sequences and sub-tasks. The objective of this conceived task was to retrieve a small vial, insert it into a rack and push them all together onto a conveyor belt. Within this sequential workflow, further sub-tasks were derived, while ensuring their interdependence. The derived tasks in the environment are commonly encountered in robotic manipulation and physics-based interactions (Billard and Kragic, 2019). With a total of seven experts (as seen in Figure 5.1), a corresponding seven sequence activation cases were conceptualised, referred to hereinafter as ***Scenarios***. Consequently, the numbering of scenarios by extend also indicates how many experts are involved; corresponding to

the number of correct sequential macro-actions necessitated for the MN to activate and orchestrate. Thereby, each scenario requires adding an additional expert, granting the overall tasks more complex. The episode termination was triggered either (i) once the button next to the conveyor belt was pushed, (ii) the maximum step count for the episode was reached or (iii) the end-effector spatially deviated too far from the centre of the scene.

5.2.6 Individual Expert Network Characteristics

Part of the goal of this work when deriving each expert NN for the hierarchical formation of ROMAN was to derive as many fundamental manipulation primitives in 3D space as possible, henceforth referred to as ***experts***. This allowed, in turn, the validation of the robustness of the architecture against increased complexity, uncertainty and dimensionality. When these elements are thereafter fused and combined together and in virtue of ROMAN's HHL, the employed task-decomposition and the ability of the MN to infer higher-level scene understandings, intricate and long horizon sequences can be realised and solved. During the derivation of these experts, it was deemed important that these should not too closely interrelate with each other, to promote increased specialising skill diversity and introduce as many manipulation skill primitives as possible. The derivation of such diverse skills will thereafter aid in the overall and increased generalisation capabilities of ROMAN towards solving intricate sequential manipulation tasks challenged by long horizon sequences (Triantafyllidis et al., 2021; Ortenzi et al., 2019). The total number of expert NNs that were trained was seven and are visually depicted in Figure 5.1. These expert NNs are also listed in full detail in Table 5.2. Table 5.3 depicts all of the integrated components of ROMAN including the architecture overview.

Action Space – The experts detailed in Table 5.2 shared identical actions. These actions included full end-effector velocities ($\alpha_1 : \pm v_x$, $\alpha_2 : \pm v_y$, $\alpha_3 : \pm v_z$), as well as controlling the binary state of the gripper ($\alpha_4 : f(\pm x_g)$). It was deemed important for the expert NNs to share the same action space to highlight the value of the proposed hierarchical framework. This is due to expert specialisation not being aided by having constrained the action space available to each expert to those that are only relevant for their respective specialisation (Triantafyllidis et al., 2023b).

State Space – In contrast to the action space, the state space for each expert differed depending on the specialising manipulation skill of each one. Consequently, only the

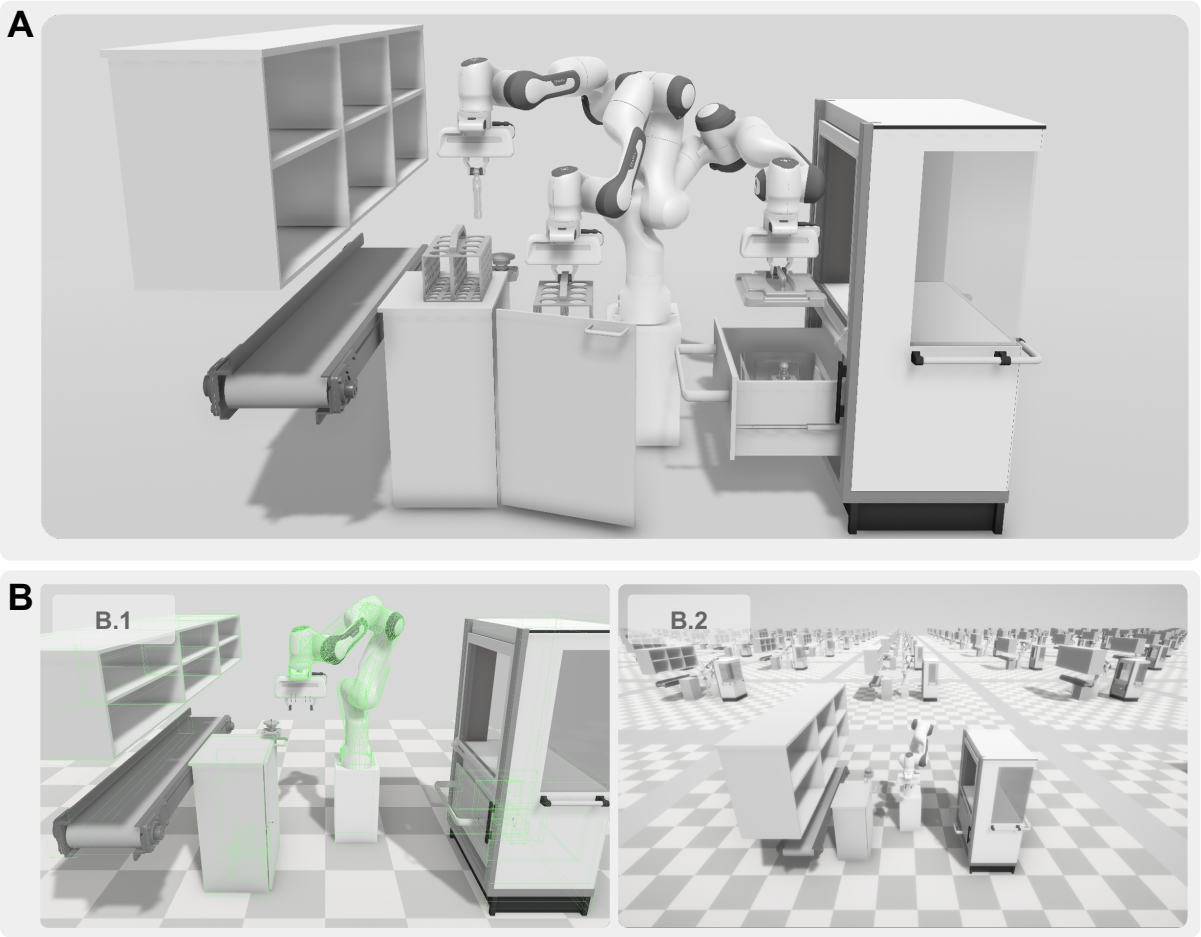

Figure 5.3: **The simulation setup of ROMAN.** The figure depicts the simulation, showcasing ROMAN and in particular the task in closer detail. **Figure A:** The simulation environment of ROMAN visualised with distinct key snapshots of its operation and in particular all phases concerned with **picking**. The figure serves as a brief visual overview of the task. **Figure B.1:** The rigid bodies and colliders highlighted in green, of both the Franka Emikra Robot, served as the main robot to test ROMAN on, including the surrounding environment. These colliders, alongside the physics configurations, presented themselves as a promising validation environment entailing realistic physics-based interactions. **Figure B.2:** The illustration of the parallelism of training ROMAN, which allowed for a notably reduced real wall-clock training times by virtue of utilising concurrent instances of the learning scene.

Table 5.2: **The summary of all experts, or skills, included in the hierarchical formation of ROMAN.** These experts are diverse in their nature and specialise in fundamental types of manipulation and interaction skills commonly seen in the daily activities of humans and in particular in robotic teleoperation.

Expert	**Symbol**	**Brief Description and Objective**
Pull-Open	π_{Pull}	Expert responsible for pulling and opening a linearly moving object, such as a common sliding drawer.
Pick and Drop	$\pi_{PickDrop}$	Expert responsible for picking and dropping an object. Similar to other *Pick* experts with a distinct difference of disregarding height offset during *placing*, hence the *dropping*. A common daily example includes when removing the lid from an object or box to expose its contents.
Rotate Open	$\pi_{RotateOpen}$	Expert responsible for rotating a door or cabinet. A common scenario when opening a cabinet or door by rotating a handle which as a whole is configured around a single rotating axis.
Pick and Place	$\pi_{PickPlace}$	Expert responsible for picking and placing an object. Similar to the *Pick and Drop* with the distinct difference of *placing* the object carefully by minimising the height drop to as close to minimal as possible. A common example includes placing a glass of water on a table.
Pick and Insert	$\pi_{PickInsert}$	Expert responsible for picking and inserting an object. Almost identical to the *Pick and Place* expert, however, with a significantly higher level of precision. The difference, as with all *Pick* experts, is the last phase of the manipulation goal and in this instance, it includes *docking* to the target location. Commonly seen when inserting a key into a keyhole and peg-insertion tasks.
Push	π_{Push}	Expert responsible for pushing an object over a surface. One of the most common and diverse examples of daily activities.
Push-Button	π_{Button}	Expert responsible for pushing and activating a human-made switch or button.

relevant information from the environment for the successful completion of the individual task was provided to the experts. This led to each NN to only focus on its individual core exteroceptive information relevant to its sub-task, and omit non-relevant ones. This decision also stemmed from the literature review of this thesis, whereby even from a neuro-scientific perspective, during the human decision-making process, the relevance of the information from the environment is determined and specified during a motor task (Wolpert et al., 2011). Hence, for a respective motor skill within a hierarchical formation, the omission of non-relevant information that is not "crucial" to the end-goal may actually aid intelligent agents while simultaneously drawing inspiration from a biological perspective. Consequently, in regards to the reward design for each of the involved NN experts, this was specific to their sub-task and the nature of their objective. These were a combination of sparse rewards, with a terminal reward at the end of a successful episode. The state and action space, including the demonstration settings and training times for each expert and the MN are detailed in Table 5.3.

The Importance of Avoiding Excessive Task-Decomposition

From Table 5.2, it can be inferred that the derived experts represent high-level abstract manipulation tasks that are fundamentally part of the daily activities and motor tasks exhibited by humans. Contrary to related work, most of the derived manipulation experts within the ROMAN are not decomposed into their most basic sub-tasks as with the works of Zhang et al. (2018b); Marzari et al. (2021). For example, in the work of Marzari et al. (2021) *Picking and Placing* is represented as a three-expert hierarchical architecture decomposed into (i) *Approaching*, (ii) *Manipulating* and lastly (iii) *Retracting*. The derived framework in this work shows that by virtue of the hybrid HHL approach in the architecture of ROMAN, *Picking and Placing* can be represented as a **single** high-level expert instead of a hierarchical composition of lower-level experts as with Marzari et al. (2021). While decomposing tasks such as *Picking and Placing* into their most primitives as with Marzari et al. (2021) allows for the derivation of more abstract cases, it does limit the potential of a high-level hierarchical model. The author of this thesis ultimately shows that ROMAN's higher-level operating experts aid the overall hierarchy and in particular the MN in addressing notably intricate tasks.

More specifically, as it will be observed further along this chapter, the higher-level operation of the experts allows the MN to focus almost exclusively on higher-level scene understandings, specifying the correct sequential activation of the individual NN com-

ponents (experts as macro-actions) and ultimately solve the intricate and long-horizon sequential tasks. In contrast, if the experts were mostly concerned with very low-level manipulation tasks, the potential of the MN would likely be limited to decomposing a rather simplistic task such as *Picking and Placing*, which could instead be the result of a single expert. Ultimately, this would likely have led to the MN being concerned with individual tasks and shifting its focus from the more important end-goal of solving the intricate sequence and long-horizon task. This is where ROMAN's core HHL architecture in addition to the derivation of its experts specialising in the highest possible level of manipulation tasks overcome these shortcomings and are able to solve notably intricate long-horizon tasks.

Furthermore, each single task-level expert NN trained via the employed hybrid learning procedure, has its own inherent robustness in facing new states during the exploration process from the RL paradigm. This allowed the MN to be trained more effectively in solving highly complex sequences over long time horizons, without the need to learn how to re-combine primitive action-based experts to achieve a given sub-task.

The Implementation of a "Dead Zone"

Specific to the evaluation of the ROMAN framework and in particular during the use of a robotic gripper, an interesting finding was observed in the preliminary testing of each expert NN skill. From the preliminary results, it was revealed that when the MN was switching between the different experts within the hierarchy, there was the possibility of dropping the object to the floor when suddenly switching from any expert concerned with *Picking* to another. An appropriate mitigation strategy for this purpose was implemented which is referred to as the **"Dead Zone (DZ)"**. More specifically, for the action of opening or closing the gripper, a range or, "zone", was implemented where the action output of $\in (-1, 1)$ would only open or close the gripper when the value reached a specified value, while the remaining range of values would essentially be ignored. Mathematically the DZ is formulated as:

$$\mathrm{DZ}(x_g) = \begin{cases} \textit{close} & \text{if } x_g \in [-1.0, -0.9], \\ \textit{remain the same} & \text{if } x_g \in (-0.9, 0.9), \\ \textit{open} & \text{if } x_g \in [0.9, 1.0]. \end{cases} \tag{5.4}$$

From the above, it can be inferred that the DZ ($\in (-1, 1)$) implementation would ignore all output values for the gripper action x_g between -0.9 and 0.9 and would only switch between

opening and closing the gripper outside these values. The aforementioned implementation led to a notable increase in overall success rates and was consequently deemed to be necessary. While this is rather task-specific, in a binary open and close case as with the incorporated robotic gripper in this work, it appears to be of value. Nevertheless, future work will ultimately determine its necessity.

5.2.7 Manipulation Network Characteristics

In this section, one of the core components of the ROMAN framework is discussed, the manipulation network. The MN acts as the master control policy, assigning and distributing weights ($\in (0,1)$) to the expert NNs it oversees and supervises. These expert NNs are treated as macro-actions in this context. The final output of those weights is defined as the sum of those weighted actions:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_i \cdot w_j \tag{5.5}$$

where the number ($m = 4$) of all actions (α_i) corresponding to each expert is controlled by a set of weights (w_j) for a given total number ($n = 7$) of experts in the hierarchy.

Similarly to the previous section, during the preliminary investigation, one of the main issues that arose when assigning the weights to the experts was the potential that the sum of all of the weights exceeded one. This ultimately led to undesired behaviour with the most notable example being that torques and forces were added beyond the robot's capabilities.

To overcome this problem, it was necessary to normalise the sum of weights assigned by the MN when activating the distinct experts it supervised. For this purpose, a normalised exponential function, also known as **softmax** was chosen. The softmax provides, in turn, a probability distribution to better isolate the expert activation during long sequential tasks. Mathematically, this is represented as:

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \qquad \text{for } i = 1,\ldots,K \text{ and } \mathbf{z} = (z_1,\ldots,z_K) \in \mathbb{R}^K. \tag{5.6}$$

whereby σ is the softmax function, z is the input vector and represented as a function of e^{z_i} denoting the standard exponential for each input, divided by the sum of all inputs K. With ROMAN, the input vector is represented as a weight vector, with each element representing the weight of every single expert, with a sum equal to $K = 7$, representing all seven distinct expert NNs.

Contrarily to the expert NNs, the state space of the MN contains the union of all of the observation spaces of each individual specialising expert. As such, while the observation space of each expert was specific to the relevant sub-task, the MN observed the entirety of the relevant sub-tasks to better distinguish which expert should be activated at which time step.

Lastly, the reward signals provided to the MN were sparse. More specifically, a reward to the MN was provided after successfully completing each sub-task and a terminal reward for successfully completing the entire sequence of tasks successfully. Figure 5.2 depicts the overall architecture of the ROMAN framework, highlighting the MN as the primary gating mechanism which centrally governs the control policy within the HHL control framework. The hyperparameter settings of the expert NNs and the MN are detailed in the Appendices, in particular Table A.11. Table 5.3 details the individual NN architecture of each component of ROMAN including the action and state space.

Table 5.3: **The overview of the hierarchical architecture and formation of ROMAN.** The table summarises ROMAN's overall NN architecture and every included component within the hierarchical framework. The table also depicts the state and action space, as well as the demonstrations provided for each NN within the ROMAN framework.

Network Architecture, Characteristics and Demonstration Settings							
MN	**Experts NNs**						
	Push-Button	Push	Pick & Insert	Pick & Place	Rotate Open	Pick & Drop	Pull Open
State Space (Vector Size)							
Total: 29	***Total: 11***	***Total: 14***	***Total: 14***	***Total: 14***	***Total: 11***	***Total: 14***	***Total: 11***
Agent Position (3)	Agent Position (3)	Agent Position (3)	Agent Position (3)	Agent Position (3)	Agent Position (3)	Agent Position (3)	Agent Position (3)
Agent Velocity (3)	Agent Velocity (3)	Agent Velocity (3)	Agent Velocity (3)	Agent Velocity (3)	Agent Velocity (3)	Agent Velocity (3)	Agent Velocity (3)
Gripper Force (2)	Gripper Force (2)	Gripper Force (2)	Gripper Force (2)	Gripper Force (2)	Gripper Force (2)	Gripper Force (2)	Gripper Force (2)
Full Environment (21)	Button Position (3)	Rack Position (3)	Rack Position (3)	Rack Position (3)	Cabinet Position (3)	Box Position (3)	Drawer Position (3)
		Conveyor Position (3)	Vial Position (3)	Rack Target (3)		Unbox Target (3)	
Action Space (Vector Size)							
Total: 7	***Total: 4***	***Total: 4***	***Total: 4***	***Total: 4***	***Total: 4***	***Total: 4***	***Total: 4***
Agent Weights (7)	Agent Velocity (3)	Agent Velocity (3)	Agent Velocity (3)	Agent Velocity (3)	Agent Velocity (3)	Agent Velocity (3)	Agent Velocity (3)
	Gripper State (1)	Gripper State (1)	Gripper State (1)	Gripper State (1)	Gripper State (1)	Gripper State (1)	Gripper State (1)
Demonstration Settings and Training Times (Number, Demo Time, Train Time)							
N = 42	N = 20	N = 20	N = 20	N = 20	N = 20	N = 20	N = 20
$t_{demo} \approx$ 42min	$t_{demo} \approx$ 7min	$t_{demo} \approx$ 6min	$t_{demo} \approx$ 12min	$t_{demo} \approx$ 10min	$t_{demo} \approx$ 7min	$t_{demo} \approx$ 9min	$t_{demo} \approx$ 7min
t_{train} = 11h 22min	t_{train} = 3h 1min	t_{train} = 3h 59min	t_{train} = 23h 30min	t_{train} = 11h 46min	t_{train} = 2h 39min	t_{train} = 3h 43min	t_{train} = 3h 18min

5.3 Results

In this section, the results as part of the evaluation of the ROMAN framework are presented. As aforementioned ROMAN is composed of a modular hybrid hierarchical architecture that is able to combine adaptive motor skills and subsequently capable of solving intricate manipulation tasks. The hierarchical architecture is supervised and activated by a central gating network referred to as the MN. The MN, in turn, is responsible for higher-level scene understandings and inferring the correct sequential activations of the incorporated specialising experts.

Baseline Evaluation – Part of the first objective was to evaluate the scalability of a hierarchical learning architecture versus a single NN approach. This allowed, in turn, to investigate whether a hierarchical formation was deemed of value and whether a monolithic NN could by itself converge towards a stable solution given complex long-horizon sequences. As such, the hierarchical formation of ROMAN was compared against a monolithic NN equivalent, sharing an identical hybrid learning procedure, NN settings as well as hyperparameters. The initial baseline evaluations between hierarchical and monolithic equivalents are presented in Section 5.3.1. These baseline evaluations are composed of two main parts, (i) with ROMAN's preliminary version consisting of five in total experts on a 2D planar setup and (ii) its final 3D version composed of seven in total experts.

Robustness, Learning and Demonstration Evaluation – Thereafter, all subsequent results are in ROMAN's final version composed of seven experts in full 3D space. In particular, subsequent sections evaluate ROMAN's robustness and versatility given: (i) different levels of exteroceptive uncertainty (Section 5.3.2), (ii) different learning paradigms via an ablation study on the internal hybrid learning procedure (Section 5.3.3) as well as (iii) the effects of a different number of demonstrations for the dataset provided to the framework (Section 5.3.4). To retain consistency and conduct a fair comparison all subsequent results presented were conducted with identical network settings (states, actions and rewards), number of demonstrations and hyperparameter values.

Adaptive Behaviour – In Section 5.3.5 the importance of ROMAN's hybrid learning procedure is highlighted, by exhibiting beyond demonstrated behaviour with the most prominent results showing dynamic adaptability to beyond demonstrated behaviour and recovery capabilities when stuck in local minima. Finally, in Section 5.3.6 the employment

of a t-Distributed Stochastic Neighbor Embedding (t-SNE) is performed on the observational vector of ROMAN's gating network (MN), to further and more comprehensively investigate its ability to differentiate between the different sequential cases.

Experimental Evaluation Approach – All subsequent results stemming from the experiments were conducted with identical network settings (states, actions and rewards), number of demonstrations and hyperparameter values to retain consistency. More details on the hyperparameters and dimensions of the networks can be found in the Appendices and in particular Table A.10, Table A.11, Table A.12 and Table A.13. Information regarding the demonstration settings and technical elaborations on the acquisition of those can be found in earlier parts of this chapter and in particular in the Methodology (Section 5.2).

Main Findings – The subsequent evaluation shows that by virtue of ROMAN's hierarchical task decomposition and the employed hybrid learning procedure, ROMAN exhibits robustness against (i) large exteroceptive observation noise, (ii) complex non-interrelated compositional sub-tasks, (iii) long time-horizon sequential tasks, as well as (iv) adaptation to cases not encountered during the demonstrated sequences. Most importantly, it is observed that by virtue of the hybrid training, exploration and exploitation are balanced leading to beyond demonstrated behaviour reflected on ROMAN's ability to dynamically coordinate the specialising experts to recover from local minima successfully while also exhibiting recovery capabilities. It is worthwhile to point out that such behaviour was not demonstrated, but rather visited by virtue of the employed hybrid learning procedure, balancing exploitation and exploration. Consequently, these findings highlight and imply the versatility and adaptability of ROMAN, enabling its usage for autonomous manipulation in intricate settings, necessitating adaptive motor skills.

Definition of Success Rate – First and foremost, it is important to define the term success rate which is hereinafter used as the primary metric validating ROMAN. In the simulated medical laboratory setting whereby ROMAN was validated, success was attained when all seven sub-task goals depicted in Figure 5.1 were satisfied. Table 5.1 details the main task, with the relevant sub-tasks and their inter-dependencies for attaining the end-goal. As such, to consider a scenario successful, all inter-related sub-tasks needed to be sequentially completed within the time limit set. In layman's terms, a single failure within a given sequence would render the whole sequence an error.

5.3.1 Limitations of Monolithic Neural Networks

As part of the preliminary investigation of hierarchical models, ROMAN was initially derived in its first version operating with five in total experts in 2D space. Thereafter, ROMAN was scaled to a notably more intricate architecture composed of seven experts, operating in full 3D space.

In both instances, it was deemed crucial to initially investigate whether a hierarchical architecture was of necessity. The rationale behind this evaluation was that if a monolithic NN is unable to render equivalent or better performance than a hierarchical model in a simple 2D setting, then extending it further to 3D would render even worse performance due to the increased dimensionality. Consequently, in this section ROMAN's preliminary and final stages are compared against two monolithic NNs with an equivalent hybrid learning procedure (visually depicted in Figure 5.2.a) for 2D and 3D respectively. These baseline evaluations allowed for a direct comparison of a monolithic versus a hierarchical approach and underlined the value of the latter. Subsequently, it allowed for the evaluation and demonstration of the advantages of a hierarchical task decomposition sharing an identical learning procedure. In this way, the value of using a hierarchical approach can be tested and verified. To retain consistency across the evaluation, the monolithic NNs had identical states to ROMAN's MN and identical actions to ROMAN's experts. To furthermore conduct a fair comparison, a total of $N = 100$ and $N = 140$ demonstrations were provided to the single NNs, accounting for ROMAN's 2D and 3D cases composed of five and seven experts pre-trained with $N = 20$ demonstrations, such that $5 \cdot 20 = 100$ and $7 \cdot 20 = 140$, respectively.

ROMAN Composed of Five Experts vs Monolithic NN in 2D – The results of the initial baseline evaluation are listed in Table 5.4. From Table 5.4 it is inferred that a monolithic NN can accommodate the least complex and with shorter time-horizon sequential scenario cases S1 and S2 very well. With S3 results show that relatively high success is still attainable, although with approximately a $\approx 15\%$ drop from S2. However, extending the horizon and complexity further as seen with S4 and S5, a monolithic NN exhibits success rates of less than $\approx 60\%$, showing significantly higher errors than with simpler and shorter sequential horizons. It is worthwhile to point out that the sudden drop in success rates observed between S2 and case S3, can to a great extent be attributed to the introduction of the **Picking and Insert** expert. Picking and Insert is arguably the most demanding expert NN as it in essence combines three low-level sub-tasks composed

of (i) *reaching*, (ii) *grasping*, (iii) *placing or inserting* similarly to Marzari et al. (2021). From the overall findings at hand, it can hence be concluded that a monolithic NN, even in a simple 2D planar setting and sharing an identical hybrid learning procedure as with ROMAN's architecture, is unable to solve longer horizon (S3, S4 and S5) and by extent time demanding sequential tasks. This, in turn, underlines the value of ROMAN's hierarchical formation and its suitability for longer and more intricate in-sequence manipulation tasks.

Table 5.4: **The summary of the success rates for the preliminary version of ROMAN operating in 2D space with 5 experts versus a single NN.** This table, illustrates the preliminary version of ROMAN in planar 2D coordinates with 5 derived experts/sub-tasks. The framework is compared against a single NN equivalent. Evaluated at an uncertainty level of $\sigma = \pm 0.5$cm.

[2-D] Preliminary Version of a Monolithic NN vs ROMAN on Case Scenarios					
$\sigma = \pm 0.5$cm	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**
	Push	*+Lift*	*+Pick & Insert*	*+Pick & Drop*	*+Pull*
Single NN	0.998	0.841	0.699	0.591	0.565
ROMAN	0.993	0.995	0.982	0.971	0.974

Table Notes: *A total of $N = 20$ demonstrations were provided to each expert. The demonstration time corresponded to $\approx$ 5-7min demonstration time, dependent on the individual expert and its complexity. A total of $N = 35$ demonstrations were provided to the manipulation network, corresponding to $\approx$ 20min demonstration time. To conduct a fair comparison, a total of $N = 100$ demonstrations were provided to the single NN in 2D to account for the five in total experts within ROMAN's preliminary hierarchical architecture version in 2D, each pre-trained with $N = 20$ demonstrations, such that $5 \cdot 20 = 100$.*
Results stem from 1,000 trials for each individual coloured cell listed in the table.

ROMAN Composed of Seven Experts vs Monolithic NN in 3D – For the next evaluation all subsequent hereinafter results for ROMAN are scaled up to 3D space with seven in total experts. This decision was made to closely approximate a realistic robotic environment, while simultaneously significantly increasing the dimensionality of the problem (Triantafyllidis et al., 2023b). Similarly to the 2D case of ROMAN and even though in a simple setting a monolithic NN was unable to converge to the success rates of ROMAN's hierarchical architecture, it was deemed nonetheless important to validate whether a monolithic NN is capable of solving sequences requiring up to seven experts in 3D space. As with the previous evaluation, an identical hybrid learning procedure was employed for the monolithic NN as with ROMAN's hierarchical architecture to retain consistency and conduct a fair comparison. From the results detailed in Table 5.5 and similarly to Table 5.4, it can be once again observed that a monolithic NN is unable to solve the intricate nature and long sequential task of the validated manipulation scenario given the same training procedure. In particular, while a monolithic NN achieves comparable

success rates to ROMAN's hierarchical formation for S1 and S2, a significant drop of approximately $\approx 30\%$ is observed from S2 to S3. For S4 and S5, a highly significant drop in success rates is observed, with success rates being less than $\approx 5\%$. Lastly, it is shown that for S6 and S7 which are arguably the most demanding cases due to their long horizons, it is observed that a monolithic NN exhibits complete failure. As part of this evaluation, extending the noise levels beyond $\sigma = \pm 0.5$ cm was disregarded due to the already notably lower success rates of the monolithic NN at such noise levels. From this experiment, it can be concluded that both in 2D and especially in 3D space, a monolithic NN approach is insufficient, highlighting the necessity of the hierarchical model of ROMAN. For a more comprehensive overview of ROMAN's and the monolithic NN's sub-task related success rates within the full sequence activations, please consult Table A.2 and Table A.3 for the 2D and 3D cases respectively. For more details and expansion regarding the monolithic NNs, including their architecture and hyperparameters, please consult the Appendices and in particular Table A.12 and Table A.13.

Table 5.5: **The summary of the success rates of ROMAN vs a monolithic NN in full 3D space with seven in total derived experts.** The full 3D version of ROMAN with the most derived sub-tasks is compared against a monolithic NN equivalent. Evaluated at an uncertainty level of $\sigma = \pm 0.5$cm.

Monolithic NN vs ROMAN on Case Scenarios							
	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**	**Scenario 7**
$\sigma = \pm 0.5$cm	*One Expert (Push-Button)*	*Two Experts (+ Push)*	*Three Experts (+ Pick & Insert)*	*Four Experts (+ Pick & Place)*	*Five Experts (+ Rotate Open)*	*Six Experts (+ Pick & Drop)*	*Seven Experts (+ Pull-Open)*
Single NN	0.997	0.981	0.583	0.032	0.028	0.000	0.000
ROMAN	0.973	0.975	0.817	0.959	0.852	0.960	0.952

Table Notes: *A total of N=20 demonstration were provided for each expert. The demonstration time was ≈ 7 to 12min, depending on the expert and its complexity. A total of N=42 demonstrations corresponding to ≈ 42min demonstration time, for the manipulation network were subsequently provided. To conduct a fair comparison, a total of $N = 140$ demonstrations were provided to the single NN to account for the seven in total experts within ROMAN's hierarchical architecture, each pre-trained with $N = 20$ demonstrations, such that $7 \cdot 20 = 140$. Results stem from 1,000 trials for each individual coloured cell listed in the table.*

5.3.2 Evaluation Against Exteroceptive Uncertainty

To this point, ROMAN's preliminary version was benchmarked entailing five experts operating in a 2D planar setting. Hereinafter, ROMAN is scaled to full 3D space with seven in total experts. This will allow the framework to not only cover and generalise to a greater extent towards the domains of robotics and simulation training, but also toward commonly seen types of daily motor activities humans regularly undertake (Triantafyllidis et al., 2021; Ortenzi et al., 2019).

Part of the first investigation in full 3D space was to evaluate ROMAN against increasing uncertainty for its exteroceptive state. The rationale behind this decision stemmed from real-life cases, whereby any physical sensor introduces certain levels of noise perturbations and one cannot assume an error-free reading (Yang et al., 2020). Consequently, ROMAN was evaluated against different levels of Gaussian distributed noise and on an external simulated vision system to approximate a Sim2Real setting.

Gaussian Distributed Exteroceptive Uncertainty

The first objective as part of evaluating ROMAN, was to test for its resilience and robustness against simulated Gaussian distributed noise as seen in real life (Yang et al., 2020). As aforementioned, a vision-based detection system was derived to approximate a realistic case of providing exteroceptive information. However, prior to proceeding to such a system, the framework is initially evaluated beyond the exteroceptive uncertainties the vision system alone would yield. Prioritising the evaluation first on Gaussian distributed noise, allowed the author to gain a better understanding and ultimately evaluate the framework's robustness to the vision system.

Consequently, different levels of noise were added, via a Gaussian distribution, in all of the exteroceptive states corresponding to positional observations of all NN within the ROMAN hierarchy. First and foremost, each expert NN of the hierarchy is evaluated for each level of noise to validate their individual robustness. Attaining robust performance of all expert NNs individually was considered imperative for further evaluating the performance of the MN. The reasoning behind this decision was to first study the individual performance of each expert in order to understand whether failures were attributed to the experts themselves or due to incorrect sequential activation by MN. This in turn minimised the covariance between the success rates of each expert and that of the MN, allowing for a more comprehensive evaluation of the latter network thereafter.

Evaluation of the Specialising Expert NNs – From Table 5.6 it can be inferred that all individual experts within the ROMAN framework, even when presented with the highest evaluated levels of noise, are robust against the tested levels of uncertainty. Another important observation that was identified was that all *Picking & -Dropping, -Placing* and especially *-Inserting* experts, were arguably the most complex ones. Indeed, observing Table 5.6, it can be seen that these three experts are more prone to errors than others when presented with higher noise levels. These results can to a great extent be

attributed to their higher level complexity as well as time horizon with these experts, which is in line with Marzari et al. (2021); Antotsiou et al. (2022).

Evaluation of the Manipulation Network (MN) – After evaluating all individual expert NNs and observing that all attain high levels of success rates, the evaluation proceeded with the MN and in particular its ability in coordinating the different expert activations in the hierarchy of ROMAN. From the seven derived experts, a corresponding seven different randomised case scenarios were derived with each scenario corresponding to the addition of another required expert. Hence, with each scenario, the overall task became all the more complex with notably higher time horizons. Table 5.1 details the derived case scenarios with their dependencies to other tasks and the evaluated sequence of manipulation tasks. Table 5.6 presents the results regarding the different noise levels and the performance of the MN. From Table 5.6, it can be inferred that ROMAN and in particular the MN is skillfully capable of activating the correct sequential activation of the expert, exhibiting robust performance even in increasingly higher levels of uncertainty. As one can infer, even though the addition of more experts notably increases the dimensionality of the problem, results show that the MN is still sufficiently resilient in the most intricate settings in S6 and S7. It is worthwhile to point out the slight performance drop observed between S3, S4 and S5 compared to S6 and S7. Even though S6 and S7 are of higher complexity and time horizons, it can be seen that in some cases, the success rates drop more in S3, s4 and S5. This discrepancy in the results is analysed and discussed further in the Results section. Further expansion on the results and in particular for a comprehensive overview of the individual sub-task success rates of the MN within the full sequence of actions can be found in the Appendices and in particular Table A.1.

Vision Based Exteroceptive Uncertainty

Having evaluated ROMAN against simulated exteroceptive noise, the next objective was to validate the robustness of ROMAN against uncertainties stemming from vision-based detection. Consequently, a simulated vision system was implemented in Unity3D using an RGB camera. It is worthwhile to point out that ROMAN, including its experts and the MN, were **not** trained on the vision system. These NNs were rather evaluated and directly tested on the vision system, to investigate the framework's robustness to a never-before-seen system closer to reality. For more details please consult Section 5.2.3, elaborating on the technical implementation of the vision detection system.

Table 5.6: **The summary of the results based on increasing levels of Gaussian distributed noise on the exteroceptive states of the expert NNs and the MN, in addition to those naturally occurring from the vision system.** The table details the success rates for all individual experts as well as that of the gating network (MN) within ROMAN for the 3D setting, across all case scenarios, based on increased levels of Gaussian noise in the exteroceptive position observations. Furthermore, the success rates stemming from the vision system are depicted where ROMAN was evaluated upon. ROMAN was not trained on the vision system.

Individual Expert Success Rates [10,000 Trials per Cell]							
Success (%)	**Push-Button**	**Push**	**Pick & Insert**	**Pick & Place**	**Rotate Open**	**Pick & Drop**	**Pull Open**
$\sigma = \pm 0.0$ [cm]	0.996	0.998	0.919	0.986	0.999	0.997	0.970
$\sigma = \pm 0.5$ [cm]	0.999	0.999	0.933	0.989	0.999	0.994	0.993
$\sigma = \pm 1.0$ [cm]	0.999	0.999	0.939	0.982	0.999	0.994	0.985
$\sigma = \pm 1.5$ [cm]	1.000	0.999	0.920	0.965	0.999	0.988	0.969
$\sigma = \pm 2.0$ [cm]	0.999	0.998	0.872	0.941	0.999	0.973	0.962
$\sigma = \pm 2.5$ [cm]	0.999	0.991	0.826	0.903	0.998	0.955	0.950
MN Success Rates [1,000 Trials per Cell]							
Success (%)	**Scenario 1** *One Expert (Push-Button)*	**Scenario 2** *Two Experts (+ Push)*	**Scenario 3** *Three Experts (+ Pick & Insert)*	**Scenario 4** *Four Experts (+ Pick & Place)*	**Scenario 5** *Five Experts (+ Rotate Open)*	**Scenario 6** *Six Experts (+ Pick & Drop)*	**Scenario 7** *Seven Experts (+ Pull-Open)*
$\sigma = \pm 0.0$ [cm]	0.976	0.972	0.847	0.951	0.728	0.954	0.903
$\sigma = \pm 0.5$ [cm]	0.973	0.975	0.817	0.959	0.794	0.960	0.952
$\sigma = \pm 1.0$ [cm]	0.977	0.990	0.798	0.946	0.776	0.933	0.939
$\sigma = \pm 1.5$ [cm]	0.980	0.986	0.720	0.846	0.722	0.836	0.841
$\sigma = \pm 2.0$ [cm]	0.967	0.986	0.737	0.837	0.753	0.820	0.815
$\sigma = \pm 2.5$ [cm]	0.973	0.986	0.723	0.763	0.697	0.719	0.744
MN Vision System Success Rates [100 Trials per Cell]							
Success (%)	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**	**Scenario 7**
Visual	0.97	0.96	0.82	0.83	0.79	0.51	0.72

From the results stemming from the vision system detailed in Table 5.6, it can be observed that using a pre-trained object detection module from the simulated camera attains high success rates from S1 to S7. These success rates are consistently high from a one-expert sequential task case scenario to up to a seven sequential long-time horizon task, highlighting ROMAN's robustness considering that the architecture was not trained on such a system. Despite an observed decrease in success rates as more sequences are added with increased time horizons, ROMAN is still able to robustly orchestrate the sequential activations of experts even amongst the most intricate tasks. The observed decrease in success rates in S6 and less in S7 can to some extent be attributed to the *Pick & Drop* sub-task. As identified in earlier parts of the Results, all experts concerned with *Picking* had arguably the highest complexity. Moreover, during the *Pick & Drop* phase in particular, the OIs in that phase were the most prone to visual occlusion than the rest of the sequences (consult Figure 5.1). Furthermore, the decrease in success rates can also be attributed to the similarity in the exteroceptive observations during the sequence of actions. This drop in success rates was also observed in previous parts of the Results section regarding the validation of ROMAN on simulated noise, whereby less complex cases rendered lower success rates than more complex ones. To properly investigate this, a t-SNE is conducted (consult Figure 5.5), which is discussed in detail further along the Results. Nevertheless, from these findings, it can be observed that despite a decrease in success rates in S6 and S7, the ability of ROMAN to perform consistently well on a system it was not trained upon demonstrates to a great extent its versatility, adaptation and ultimately its robustness.

5.3.3 Ablation Study on ROMAN's Learning Paradigms

For the next validation, a comparison of state-of-the-art learning paradigms is conducted and in particular, the learning algorithms that are included and fused as part of the hybrid learning approach at the core of ROMAN's architecture. The combination of different learning paradigms, including BC, RL and GAIL as well as their pairs (HHL, HRL and HIL), are compared similarly to Marzari et al. (2021); Zhang et al. (2018b). As previously stated, ROMAN utilises BC to warm-start the policy via supervised learning and from that point onward makes use of intrinsic r^i (IL: GAIL) and extrinsic r^e (RL) rewards via PPO for training. Within the fusion of these learning paradigms, ablations are conducted to the training procedure by excluding at least one of the previous paradigms. More details on the HHL architecture can be found in the Methods section. The results

of the ablation study are detailed in Table 5.7.

Table 5.7: **The results of the ablation of the different paradigms within the architecture of ROMAN.** Success rates across all seven scenarios, between different algorithmic comparisons of HRL, HIL and their combinations, used to train ROMAN.

Algorithm Comparison Within ROMAN							
	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**	**Scenario 7**
$\sigma = \pm 0.5$cm	*One Expert (Push-Button)*	*Two Experts (+ Push)*	*Three Experts (+ Pick & Insert)*	*Four Experts (+ Pick & Place)*	*Five Experts (+ Rotate Open)*	*Six Experts (+ Pick & Drop)*	*Seven Experts (+ Pull-Open)*
RL (r^e)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GAIL (r^i)	0.980	0.468	0.559	0.012	0.003	0.001	0.000
BC	0.986	0.978	0.786	0.660	0.525	0.722	0.760
RL (r^e), *GAIL* (r^i)	0.981	0.468	0.570	0.009	0.005	0.006	0.004
BC, *RL* (r^e)	0.995	0.897	0.841	0.683	0.492	0.754	0.774
ROMAN's †	0.973	0.975	0.817	0.959	0.852	0.960	0.952
	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**	**Scenario 7**
$\sigma = \pm 1.0$cm	*One Expert (Push-Button)*	*Two Experts (+ Push)*	*Three Experts (+ Pick & Insert)*	*Four Experts (+ Pick & Place)*	*Five Experts (+ Rotate Open)*	*Six Experts (+ Pick & Drop)*	*Seven Experts (+ Pull-Open)*
BC	0.995	0.990	0.712	0.573	0.474	0.563	0.632
BC, *RL* (r^e)	0.996	0.895	0.881	0.766	0.562	0.696	0.729
ROMAN's †	0.977	0.990	0.798	0.946	0.776	0.933	0.939
	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**	**Scenario 7**
$\sigma = \pm 2.0$cm	*One Expert (Push-Button)*	*Two Experts (+ Push)*	*Three Experts (+ Pick & Insert)*	*Four Experts (+ Pick & Place)*	*Five Experts (+ Rotate Open)*	*Six Experts (+ Pick & Drop)*	*Seven Experts (+ Pull-Open)*
BC	0.838	0.678	0.609	0.205	0.190	0.111	0.075
BC, *RL* (r^e)	0.947	0.841	0.725	0.442	0.363	0.246	0.100
ROMAN's †	0.967	0.986	0.737	0.837	0.753	0.820	0.815

Table Notes: *A total of N=42 demonstrations, corresponding to approximately 42min, for the manipulation network were provided for the learning approaches requiring a demonstration dataset (i.e. BC and GAIL).* ***Notes on BC:*** *Supervised learning on the demonstration dataset.* ***Notes on GAIL:*** *Use of IL intrinsic rewards (r^i) provided to PPO.* ***Notes on RL:*** *Use of task extrinsic rewards (r^e) provided to PPO.* ***ROMAN's*** †***:*** *Default HHL approach combining BC, IL (via r^i) and RL (via r^e). Tested on $\sigma = \pm 0.5$cm noise, with increasing levels to $\sigma = \pm 1.0$cm and $\sigma = \pm 2.0$cm for algorithms scoring high. Results stem from 1,000 trials for each coloured cell listed in the table.*

Initially, it can be observed that the most apparent result of all is that the exclusive use of **RL** via extrinsic rewards (r^e) exhibits complete failure across all case scenarios. This failure can to a great extent be explained by the the high complexity and long time horizon of the tasks studied, which are, in turn, effectively unattainable via random exploration of the action space. Exclusively using the intrinsic rewards (r^i) provided by **GAIL** or combining it with extrinsic (r^e) for **RL and GAIL**, revealed significantly higher success rates. These were nonetheless limited to S1 to S3, with long-horizon tasks from S4 and onward still being unattainable.

Regarding the exclusive use of **BC**, a significant boost in success rates was initially observed. More specifically, BC attained higher levels of success than the other learning paradigms across most of the derived scenario cases. However, upon closer inspection,

longer in-time horizon sequential tasks, such as S4 to S7, exhibited overall lower performance when compared to ROMAN's default learning approach. This performance degradation can to a great extent be explained by the larger variance exhibited in the trajectories visited due to the compounding of errors throughout the trajectory, which are in line with previous work (Ho and Ermon, 2016; Rajeswaran et al., 2018; Ross et al., 2011). Indeed, from the reviewed literature (Davchev et al., 2022; Marzari et al., 2021; Zhang et al., 2018b), solely depending on BC appears to yield rapid performance degradation as the task time horizon is increased and more out-of-distribution states are presented and ultimately leading to distribution drifting. As it can be observed from the results, neither the sole use of **BC** nor when coupled with extrinsic r^e rewards, **BC and RL**, attain stable performance at longer horizons, even at $\sigma = \pm 0.5$cm noise levels. To further evaluate these findings, both **BC** and **BC and RL** were further tested on increased levels of noise of $\sigma = \pm 1.0$cm and $\sigma = \pm 2.0$cm.

Increasing the noise levels at $\sigma = \pm 1.0$cm, a slight drop in success rates is observed for both **BC** and when combined with extrinsic r^e rewards, **BC and RL**. At $\sigma = \pm 1.0$cm ROMAN's default settings still attain the highest success rates even for the longest sequential tasks such as S6 and S7.

Extending the noise level to $\sigma = \pm 2.0$cm, a highly significant drop in success rates is observed for both **BC** and **BC and RL**, especially during longer sequential tasks. In particular, at this four-fold increase of uncertainty compared to $\sigma = \pm 0.5$cm, it can be observed that **BC** is effectively unable to attain stable performance. This is especially seen when dropping to less than $\approx 20\%$ in success rates from S4 to S7. For **BC and RL**, the addition of the extrinsic r^e rewards renders marginally higher success rates, yet still not to a notable degree. While indeed ROMAN's default hybrid learning approach also exhibits a slight drop in success compared to previous levels of noise, it still skillfully attains significantly higher degrees of resilience, as shown in its success rates. These findings suggest and highlight the necessity of avoiding "naively" imitating demonstrations by solely depending on BC.

Consequently, it can be inferred that the proposed HHL approach within ROMAN's architecture is notably advantageous in overcoming increased exteroceptive uncertainties and the complexities associated with longer time-horizon sequential tasks than other combinations of existing learning paradigms. In particular, the high success rates of ROMAN can be attributed to the following:

i Utilising BC until a given epoch and limited to merely warm-starting the policy optimisation;

ii Utilising the intrinsic reward (r^i) provided by GAIL's discriminator, after the cutoff point of BC, to further minimise the divergence of the agent and that of the expert human demonstrator;

iii The addition of an extrinsic reward (r^e) stemming from the RL paradigm to allow the agent to explore beyond demonstrated behaviour.

In virtue of the above learning paradigms, their appropriate fusion as well as the employed high-level task decomposition which in turn off-loads the MN from low-level supervision of skills that can be automated at the expert-level, renders ROMAN as a whole capable of achieving a balance between exploitation and exploration. This balance is detailed further along this chapter, ultimately leading ROMAN in exhibiting dynamic adaptation in the presence of non-demonstrated cases with recovery capabilities going beyond imitated behaviour. These results also highlight the shortcomings of conventional state machines for intricate tasks necessitating adaptive motor skills, underlining the value of ROMAN's architecture. Further expansion on the ablation study are listed in the Appendices and in particular Table A.4, Table A.5a and Table A.5b.

5.3.4 Evaluation of Different Data Set Sizes

As part of the final evaluation, the effects of different numbers of demonstrations were compared. These demonstrations were provided to the dataset by the expert human and were used to evaluate the MN's policy and ultimately the hierarchical architecture of ROMAN as a whole. In this evaluation, the effects of $N = 7$, $N = 21$ and $N = 42$ number of demonstrations were analysed on the success rates across all scenarios for the MN, which corresponded to approximately 7min, 25min and 42min of data acquisition in real wall-clock time respectively.

From Table 5.8, it can be inferred that even a relatively small number of demonstrations for the MN (N=21), which corresponds to only 3 demonstrations for each of the seven sub-tasks, is sufficient to render ROMAN robust and attain high successes with lower bounds of 71.80% and upper bounds of 99.40%. Doubling the number of demonstrations to N=42, which corresponds to 6 demonstrations for each of the seven sequence cases, yields higher success rates than N=21, yet not to a notable degree. Finally, a one-shot

Table 5.8: **The effects of the different number of demonstrations for ROMAN.** The table highlights the effect of the different number of demonstrations on the ROMAN framework across all seven derived sequential scenario cases.

Demonstration Comparison N=7, 21 and 42 on Case Scenarios							
	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**	**Scenario 7**
$\sigma = \pm 0.5$cm	*One Expert (Push-Button)*	*Two Experts (+ Push)*	*Three Experts (+ Pick & Insert)*	*Four Experts (+ Pick & Place)*	*Five Experts (+ Rotate Open)*	*Six Experts (+ Pick & Drop)*	*Seven Experts (+ Pull-Open)*
N = 7	0.775	0.876	0.680	0.378	0.360	0.008	0.005
N = 21	0.994	0.921	0.718	0.945	0.903	0.929	0.958
N = 42	0.973	0.975	0.817	0.959	0.852	0.960	0.952

Table Notes: *The total number is divided by the number of scenarios, with N=7, 21 and 42 i.e. denoting 1, 3 and 6 demonstrations per scenario respectively. Results stem from 1,000 trials for each coloured cell listed in the table.*

demonstration of each scenario (i.e. N=7), does not yield high success rates during more complex sequences as shown in S4 to S7. For further expansion on the success rates, at the sub-task level, stemming from the effects of the studied demonstrations for the MN please consult Table A.6 in the Appendices.

5.3.5 Dynamic Adaptation and Recovery Capabilities

In virtue of ROMAN's HHL approach and by the extent of the balance of exploration and exploitation within the architecture, a positive adaptation during non-demonstrated cases is observed. Most notably, recovery capabilities are exhibited in what would have otherwise been sequential errors. From the experiments, it was occasionally observed that experts within the hierarchy and in particular the *Pick and Place/-Drop/-Insert* experts could seldom fail in retaining a firm grasp during picking, resulting in the grasped object being dropped. It is worthwhile to point out that this occurred fairly infrequently as supported by the evaluation of each individual specialising expert and the MN's performance as detailed in Table 5.6. Upon further evaluation, it was observed that during such seldom expert-level failures, the MN started to recognise the sub-task states and steadily learned a new orchestration of expert weight assignments until it rendered the particular sub-task state successful. Consequently, this positive adaptation of the framework to initiate a re-grasping procedure and retain robustness can to a great extent be attributed to its hybrid learning approach and the subsequent balance of imitation and exploration. Some of these re-grasping procedures are visually depicted in Figure 5.4.a and Figure 5.4.b.

Another important characteristic of ROMAN was observed with the MN. From the results, it was observed that the MN exhibited the ability to recover from local minima by rapidly

weight-switching to different experts when there was the necessity to do so. In particular, during the sequential activation of all included experts within the architecture, there was a rare occurrence in which the robot gripper could get stuck under the cabinet while retrieving the rack during *Pick and Placing*. In these rare cases, the MN would recognise the state of the gripper and activate other experts, to alter the trajectory and move the gripper away from the cabinet until it was collision-free. Once the MN ensured that the gripper was collision-free, it would then finally successfully recommence the task. An example of this occurrence is visually depicted in Figure 5.4.c.

It can hence be inferred that these results highlight to a great extent the value of the proposed hierarchical architecture of ROMAN, which combines the advantages of fusing supervised (BC), imitation (GAIL) and reinforcement learning (RL) paradigms. As it was observed, the fusion and modification of the aforementioned learning algorithms and the higher-level task decomposition employed in ROMAN resulted in robust performance as demonstrated in the success rates in earlier parts of this chapter. Most importantly, by balancing the exploration and exploitation trade-off via the employed hybrid learning procedure, ROMAN showed adaptation to cases beyond demonstrated behaviour by virtue of the exploratory nature of utilising RL, ultimately exhibiting dynamic recovery capabilities.

5.3.6 Similarity of Sequences Analysis via t-SNE

Following the evaluation of ROMAN and more specifically the MN at sequential activation of the different experts in different scenarios, further evaluation was conducted to test for potential similarities and patterns. As it was observed, in most of the results, scenario cases S3, S4 and S5 exhibited lower success rates, compared to S6 and S7, even though the latter two are by definition more intricate tasks entailing longer horizons. Consequently, to evaluate this discrepancy and qualitatively study the capability of the MN to determine the necessary sequential activation of the experts throughout the long horizon task based on its observations, dimensionality reduction via the t-SNE was conducted. This allowed, in turn, to comprehensively evaluate the similarities in the observation vector of the MN across the different scenarios cases. Ultimately, this investigation would shed additional light as to how well the MN is capable of differentiating between the observation vectors and by the extent of the effects on its policy performance. The t-SNE plots are depicted in Figure 5.5.a and Figure 5.5.b.

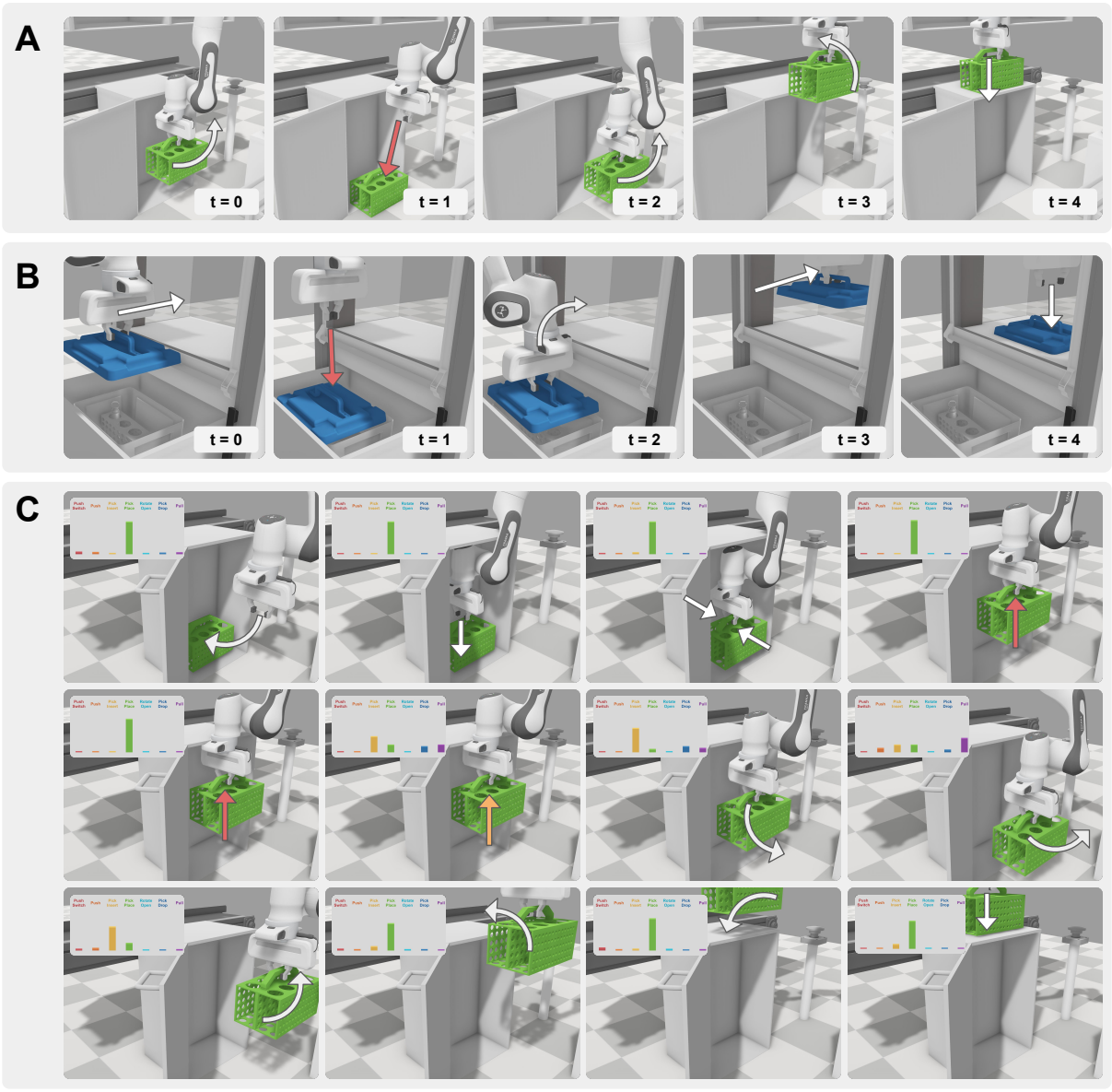

Figure 5.4: **The demonstrated ability of ROMAN to adapt to environmental demands and in particular beyond demonstrated behaviour, exhibiting dynamic recovery capabilities by balancing exploitation and exploration via the employed HHL approach. Figures (A) and (B):** The policy adaptation of ROMAN during failures during *Pick and Place* as well as *Pick and Drop* sub-tasks respectively. These intermediate failures are either attributed to an expert or a gating network (MN) error. In these fairly infrequent instances, error cases ($t = 1$) of these experts are shown. Nevertheless, by virtue of the employed hybrid learning employed throughout all NNs within ROMAN, these experts quickly re-adapt and re-grasp the items ($t = 2$ to $t = 4$) to successfully complete the sequence. These error occurrences are most notably due to a combination of incorrect grasping of objects, expert trajectories or sequential activations stemming from the supervising MN. **Figure (C):** The figure represents twelve snapshots in time with a sequence from left to right and top to bottom, with the weight assignments by the MN highlighted. The figure depicts the ability of the MN of the ROMAN architecture to dynamically adapt in cases that were not encountered in the initial demonstrations but rather visited states during the exploratory nature of RL during training as the positive result of balancing exploitation and exploration, leading to recovery capabilities from local minima.

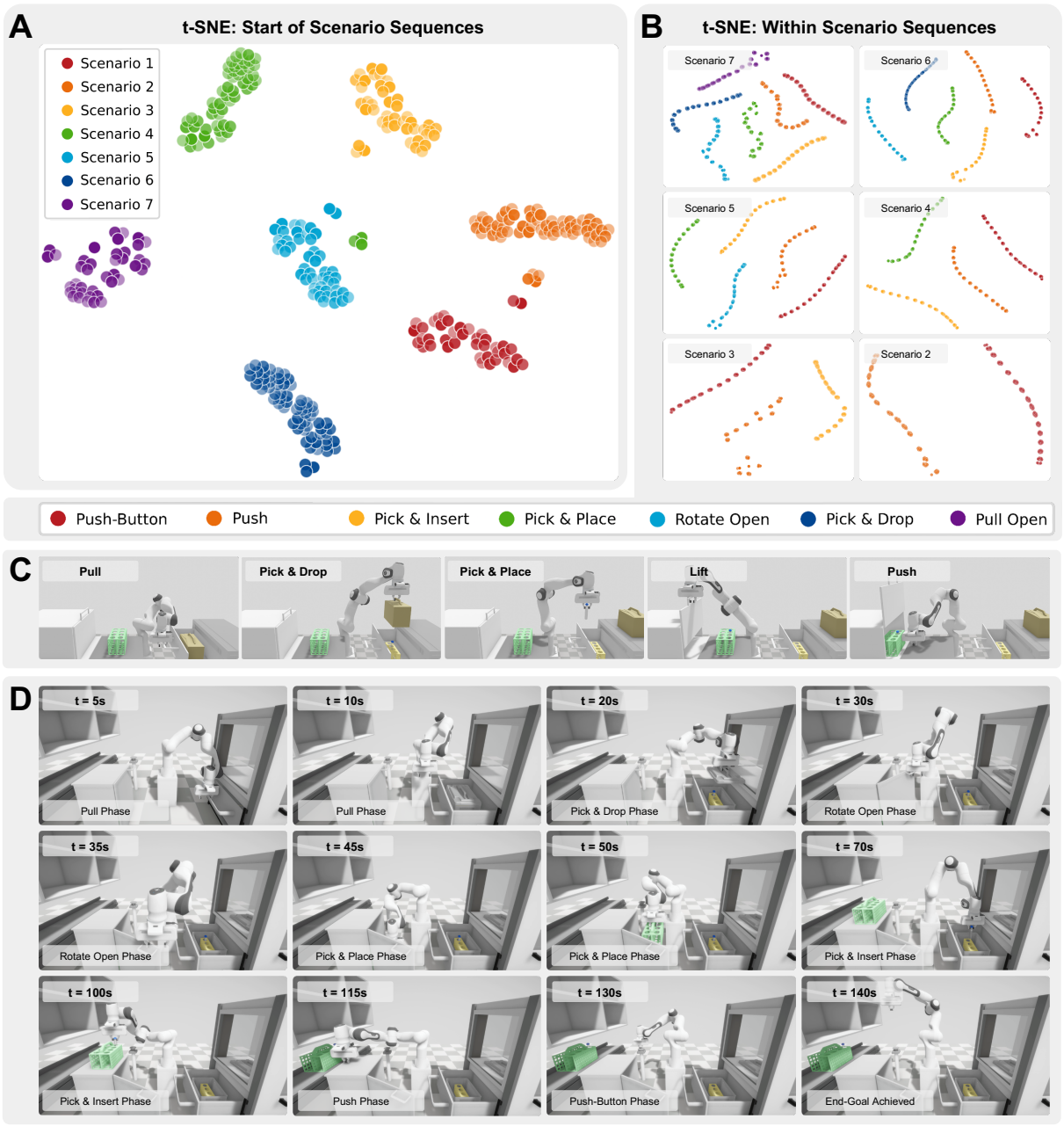

Figure 5.5: **The analysis of the MN observations using the t-Distributed Stochastic Neighbour Embedding (t-SNE), with visualised snapshots depicting ROMAN's completion of sequential tasks in 2D and 3D scenarios.** The t-SNE projects the 29-dimensional MN state vector into 2 dimensional space. Principal Component Analysis (PCA) was used for warm-starting the t-SNE projection. **Figure (A):** The depiction of the state vectors *at the start* of each of the seven case scenarios, sampled at 1,000Hz for 1s. A total of 1,000 samples were projected, with a perplexity equal to 400. **Figure (B):** The illustration of the state vectors *during* the sequence of actions contained in each case scenario. Sampling was for the first 1.5s of each expert sequence. Consequently, as these are sampled *within* the sequence of actions, they appear "trajectory-like". This is due to the robot and the objects manipulated being in motion during the sampling. A total of 1,500 samples were projected with a perplexity equal to 200. Six out of seven scenario cases are depicted, this is in practice due to S1 only including a single expert activation and is hence omitted from the analysis. **Figure (C):** ROMAN in its initial 2D stage depicting all five distinct sub-tasks managed by each expert respectively. **Figure (D):** ROMAN in its final stage in the most complex 3D setting and longest horizons with seven different specialising experts.

The **first t-SNE (in Figure 5.5.a)** that was conducted was based on the MN observations **at the beginning of each scenario** to analyse potential similarities between the MN observations corresponding to different scenario cases. As visually depicted in Figure 5.5.a, while the MN can skillfully differentiate all of the scenario cases to a great degree, cases S3, S4 and S5 present a slight overlap between each other. In particular, it would appear that the state vectors between these three cases are relatively similar, which would explain to a certain degree the difficulty of the MN to differentiate the correct sequence activation. The author of this thesis hypothesises that this is particularly the case during the initiation of each sequence whereby the starting position of the end-effector is randomised, as opposed to being the ending position of the previous sub-task, which would, in turn, lead to a slight performance drop in the success rates of those cases.

The **second t-SNE (in Figure 5.5.b)** that was conducted was based on the MN observations **of each separate activation for each scenario within** studied. Figure 5.5.b depicts the t-SNE analysis of the similarities in the MN observations throughout different expert activations within each of the seven case scenarios. From Figure 5.5.b, it can be inferred that by sampling within the sequence of actions, a low-dimensional projection of the trajectory of the MN observation vectors during the expert activations is obtained. Essentially, this is due to the shifting of the spatial states of the objects in the scene as well as the end-effector being in motion *during* the sequence of actions.

Overall, from Figure 5.5.b, no notable overlaps between the activation of the different experts within each scenario can be observed. This suggests that the MN is capable of distinctly activating the experts throughout the sub-task completion. Consequently, regarding the slight performance drop for S3, S4, and S5 compared to S6 and S7 observed in the success rate results of ROMAN, it can be concluded that the failures that account for the slight drop in performance do not occur **during** (Figure 5.5.b) the course of the sub-task completion, but rather at the **beginning** (Figure 5.5.a) of the sequences. This is supported and attributed to the randomised initialisation as observed and analysed in Figure 5.5.a.

5.4 Discussion

From the experiments and results, it was observed that in contrast to the related work (Marzari et al., 2021), by virtue of the higher-level hierarchical task decomposition com-

pared to existing hierarchical solutions, ROMAN yielded robust performance in notably more intricate sequential tasks. This higher yield in performance is due to the offloading of "unnecessary" information for the MN, which in turn operates on higher-level scene understandings and orchestration of sequences rather than focusing on low-level supervision of sub-tasks. Consequently, these results support that ROMAN is capable of orchestrating significantly more complex sequential tasks, of longer time horizons as well as higher dimensionality tasks than similar work in physics-based manipulation (Marzari et al., 2021; Rajeswaran et al., 2018; Fox et al., 2019).

Furthermore, by virtue of ROMAN's HHL architecture, detailed earlier in Section 5.2 and visually depicted in Figure 5.2, the overall architecture exhibited positive and successful adaptations to cases beyond those encountered in the demonstrated sequence, with a notable advantage being the capability of recovering from local minima. It is worthwhile to point out that these cases were not demonstrated before. These results imply that while IL is highly effective in providing a baseline via exploitation, a balance between imitating the demonstrations and maximising the extrinsic reward (r^e) stemming from the RL paradigm through random exploration is critically important for the adaptation and success in such cases. This balance between exploration and exploitation of ROMAN's architecture, also draws inspiration from a biological perspective as identified in the related works of Zador (2019); Saxe et al. (2021).

Moreover, it was observed that ROMAN's central gating network (the MN) was capable of solving the most complex and longest horizon sequential manipulation tasks skillfully. Upon further investigation, the performance drop in some of the tasks with lower complexity, such as S3, S4 and S5 compared to more complex ones such as S6 and S7, was primarily attributed to the randomised initialisation of tasks and the difficulties of the MN to differentiate, to some extent, between those states.

5.4.1 Limitations and Future Work

As part of the derivation of the ROMAN architecture in this thesis and the HHL approach within its architecture, the following limitations were identified, which, in turn, constitute future work and directions to improve upon. These include:

- *Improving State Vector Differentiation* – Part of the future work of this thesis is to explore more sensory feedback states to differentiate ambiguous cases, or provide ROMAN with some form of "memory mechanism". The latter can either be

achieved by explicitly expanding the observation space with past states or implicitly by employing a recurrent NN architecture for the MN.

- *Higher Dimensionality and Complexity* – Another important factor to consider as part of future work would be to extend the ROMAN framework to even higher-dimensionality problems. Some of these problems include (i) multi-agent hierarchical learning as well as (ii) tasks requiring bi-manual operation.
- *Use of Immersive Technologies as the Means of Input* – Moreover, leveraging immersive technologies such as VR and MR are also a promising means of providing demonstrations and rendering the execution of complex tasks for human demonstrators notably easier than conventional input devices such as the ones explored in this chapter (Triantafyllidis et al., 2021; Triantafyllidis et al., 2020).
- *"Recycling" Existing Demonstrations for Imitation* – Lastly, another potential improvement, or alternative, would be the derivation of the lowest possible level of "fundamental" manipulation skills. While this may require the use of a hierarchical framework with an increased number of levels within its hierarchy and perhaps limit the ability of the MN to focus solely on higher-level scene understandings, it may in actuality be beneficial in terms of re-using demonstrations as shown in the related work of Fox et al. (2019).

Nevertheless, from this chapter, it is supported that the biologically-inspired ROMAN architecture is capable of solving notably intricate sequential manipulation tasks entailing long horizons with generalising capabilities. In particular, from the results of this work, the robustness of ROMAN and its HHL architecture is supported against (i) exteroceptive observational uncertainty levels, (ii) the presence of many complex non-interrelated compositional sub-tasks, (iii) long-horizon sequences as well as (iv) adaptation to cases beyond demonstrated behaviour, with the notable capability of dynamically coordinating experts to recover from local minima. Consult Chapter 7 for more details regarding a comprehensive discussion surrounding ROMAN and concluding remarks.

Code Availability

The code of ROMAN is made publicly available at `https://github.com/etriantafyllidis/ROMAN`, (Triantafyllidis et al., 2023a).

Chapter 6

Intrinsic Language-Guided Exploration for Intricate Tasks

To this point, advances in sensory-motor perception have been presented and how those constituted the basis for deriving the biologically-inspired ROMAN framework.

Nevertheless, with the hopes of further solidifying the derivation of intelligent and adaptive methods (e.g. ROMAN), additional biological-mimicry may aid in aligning the extraordinary human capabilities closer to embodied intelligence. For instance, human cognition is most notably enhanced by the power of language. This appears to especially hold true surrounding the thought process, which, in turn, can elicit meaningful reasoning to address intricate settings as suggested by Lupyan and Clark (2015). Perhaps, if embodied intelligence is further to be rooted in biology beyond what was presented to this point (in particular Chapter 5), a mechanism that can harness the instructive and explanatory power of language may be necessary.

As part of the last work of this research thesis, the power of language is investigated. In particular, the utilisation of Large Language Models (LLMs) are investigated in fostering exploratory potential for intricate tasks embodied intelligence and most notably robotics may be faced with, inspired throughout from a biological perspective and with the hopes of AI mimicking the extraordinary capabilities exhibited by humans.

Parts of this chapter are currently under review as: E. Triantafyllidis, F. Christianos and Z. Li. (2023). "Intrinsic Language-Guided Exploration for Complex Long-Horizon Robotic Manipulation Tasks". In: *International Conference on Robotics and Automation (ICRA 2024)*.

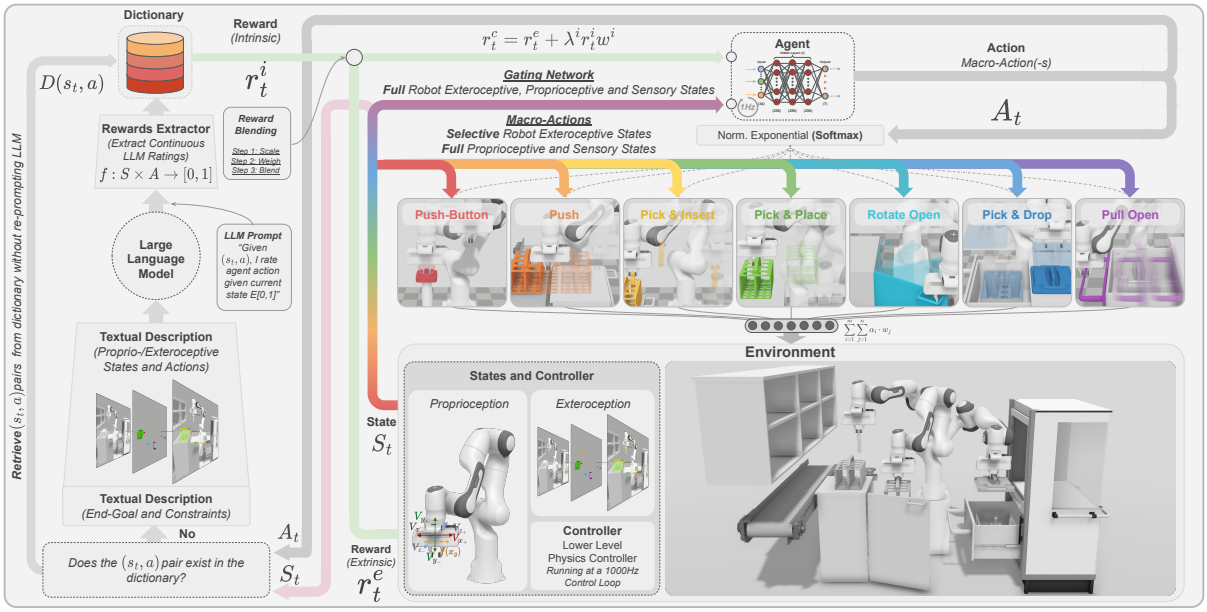

Figure 6.1: **Schematic illustrating IGE-LLMs applied on ROMAN's hierarchical architecture for solving intricate robotic manipulation tasks entailing sparse rewards and long horizons.**

As it was identified in the previous chapter (Chapter 5), a common challenge when employing learning-based approaches especially RL is the need for exploration (Schäfer et al., 2022); which holds especially true in intricate settings with sparse rewards, with the most notable being ROMAN's long horizon environment. LLMs and their context and commonsense aware reasoning provide a promising means of high-level textual interpretation of environmental demands and what actions are necessitated to meet these (Du et al., 2023; Yu et al., 2023). Nevertheless, LLMs are susceptible to occasional prompt inaccuracies (Carta et al., 2023; Brown et al., 2020), hence direct reliance on their outputs should be avoided. What if, instead of directly relying on their prompts and hence utilising these as main policy drivers, LLMs can instead be leveraged as a mere intrinsically guided proxy reward?

This chapter addresses this challenge and presents the last contribution of this thesis – the **Intrinsically Guided Exploration from Large Language Models (IGE-LLMs)** framework. By leveraging LLMs as an assistive intrinsic proxy reward, IGE-LLMs guides the exploratory process in RL to address the challenges of intricate long-horizon with sparse rewards robotic manipulation tasks, most notably ROMAN's. IGE-LLMs is evaluated on two environments, one challenged with exploration and thereafter ROMAN's intricate robotic manipulation task challenged by both exploration and long-horizons.

The results of this chapter will show that the IGE-LLMs framework exhibits notably higher performance over related intrinsic methods and overcomes the inherent limitations of the direct use of LLMs in the decision-making process. Moreover, the framework's modularity is highlighted by its ability to be combined with and complement existing learning methods. Lastly, the framework's insensitivity to different intrinsic scaling parameters is highlighted which is a common challenge when employing intrinsic rewards (Schäfer et al., 2022). Most importantly, IGE-LLMs maintain robustness both against increased levels of noise uncertainty and longer horizons over related state-of-the-art approaches. The general overview and principles of IGE-LLMs are visually shown in Figure 6.2. In a more specific context, IGE-LLMs applied on ROMAN's hierarchical architecture, extending it as a language-guided hierarchy for solving intricate robotic manipulation tasks, is shown in Figure 6.1.

Chapter Outline – This chapter is organised into a total of five sections. First, a brief overview of prior work is presented in Section 6.1, forming the basis and motivation for the derivation of IGE-LLMs. Thereafter, the methodology in Section 6.2 is outlined, whereby the technical preliminaries are presented prior to the technical elaboration of IGE-LLMs. Subsequently, details regarding the evaluation and subsequent results of the framework are presented in Section 6.3 and Section 6.4 respectively. Next, Section 6.5 presents the conclusions of this chapter and discusses the value of IGE-LLMs, framed with the state-of-the-art. Finally, Chapter 7 summarises future work and limitations of the work of this chapter.

6.1 Motivation on Language-Guided Exploration

This section provides an introduction and overview of the challenges faced by learning approaches, particularly RL, especially when applied to intricate robotic tasks. Furthermore, methods such as intrinsic rewards are mentioned and how these can complement and foster exploratory potential while highlighting their own limitations and subsequently underscoring the rationale for leveraging LLMs as a promising alternative for promoting exploration.

The Necessity for Effective Exploration in RL

An essential challenge RL approaches are faced with are intricate settings that are challenged by sparse rewards, especially prominent in long horizon robotic manipulation tasks (Davchev et al., 2022; Fox et al., 2019; Triantafyllidis et al., 2023b). In such RL settings, an essential challenge is the need for exploration. As immediate feedback from the environment may not be readily apparent and is usually in the form of sparse rewards, methods of promoting exploration are necessitated (Schäfer et al., 2022). In such settings, for agents to randomly attain a sparse goal is highly unlikely due to the exploration necessitated (Schäfer et al., 2022), which is further aggravated in long-horizon tasks such as robotic manipulation, consult Davchev et al. (2022); Triantafyllidis et al. (2023b), most notably ROMAN's environment. Ultimately, random attainment of the goal may be feasible, yet this would require prohibitively resource-intensive learning sessions (Thor and Manoonpong, 2022; Yang et al., 2020).

To overcome to some extent the challenges of exploration, methods such as ϵ-greedy policies (Watkins, 1989) or the introduction of noise in the actions of an agent (Lillicrap et al., 2016) can be inefficient and in particular during intricate robotic manipulation tasks or environments challenged by long-horizons (Triantafyllidis et al., 2023b; Schäfer et al., 2022; Pathak et al., 2017). On the other hand, while hand-crafting specific dense rewards may overcome some of the challenges of exploration, these require substantial engineering efforts. Moreover, hand-crafting such rewards notably limits generalisation to other tasks and ultimately renders these highly case dependent (Pathak et al., 2017; Du et al., 2023; Amodei et al., 2016).

Strengths and Shortcomings of Intrinsic Rewards

A promising alternative for encouraging exploration in such settings, particularly effective in sparse reward environments, is the introduction of intrinsic rewards (r^i) (Schäfer et al., 2022; Pathak et al., 2017; Flet-Berliac et al., 2021; Burda et al., 2019; Oudeyer et al., 2007; Chentanez et al., 2004). With intrinsic rewards, an additional reward (r^i) is computed and added to the existing extrinsic reward (r^e) stemming from the environment. These rewards can provide more immediate feedback to the agent, especially when the existing extrinsic reward is sparse. Without these rewards, achieving the sparse goal through random exploration becomes improbable or would require excessively and prohibitively long learning sessions (Thor and Manoonpong, 2022; Yang et al., 2020). Nevertheless, while

intrinsic rewards enhance exploration in long-horizon tasks and environments challenged by sparse extrinsic rewards, such methods can still overemphasise noisy and non-relevant to the end-goal state transitions due to prediction errors (Pathak et al., 2017; Taiga et al., 2020).

Enhancing RL Exploration with Contextual Guidance from LLMs

An alternative of providing an emphasis on more relevant to the task at hand state transitions in achieving an end goal could be by leveraging LLMs. LLMs are capable of providing useful context and common-sense aware reasoning that could aid agents in intricate settings challenged by sparse rewards and long-horizons. By emphasising more relevant state transitions given a certain end-goal and its textual description, LLMs could provide agents with more relevant to the task at hand exploratory potential (Du et al., 2023; Yu et al., 2023; Radford et al., 2019). Due to the advantages of language-based instructions stemming from LLMs, their emergence has shown promising applications in conjunction with deep learning approaches (Du et al., 2023; Yu et al., 2023; Kwon et al., 2023; Wu et al., 2023; Carta et al., 2023) and robotics (Yu et al., 2023; Ahn et al., 2022).

Unfortunately, a primary and notable limitation of LLMs is that their language outputs may not always be reliable and can occasionally be inaccurate (Carta et al., 2023; Brown et al., 2020). Hence, while LLMs provide promising means of common and context-aware reasoning; when a safety-guard mechanism is absent, their language-based outputs cannot be guaranteed (Brown et al., 2020). Moreover, recent methods leveraging LLMs are faced with notable challenges. These range from grounding and constraining LLMs (Carta et al., 2023; Ahn et al., 2022; Brown et al., 2020), limiting potential emergent behaviour by instructing the language model to follow specific manual guidelines (Wu et al., 2023), directly relying on goal suggestions that may deviate from the relevance of the task at hand (Du et al., 2023) or often require substantial prompt engineering efforts to elicit more reliable reasoning (Wei et al., 2023).

To overcome these challenges, the author of this thesis proposes the IGE-LLMs framework. IGE-LLMs leverages LLMs as an assistive intrinsic proxy reward source to complement conventional RL settings utilising the traditional extrinsic reward from the environment. The author asserts that in this process, naturally occurring inaccuracies and non-optimal replies stemming from the LLM (Carta et al., 2023; Brown et al., 2020), will not result in the policy to learn in a sub-optimal way as such replies are merely used for guidance as

an intrinsic reward which is decayed over time, with the extrinsic reward still eventually being the main policy driver.

On one hand, the author hypothesises that utilising LLMs with their context and commonsense ware reasoning can suggest and prioritise states that are directly necessary for obtaining the sparse end-goal and subsequently mitigate the limitations of current intrinsic methods plausibly exploring non-relevant to a task states (Pathak et al., 2017; Taiga et al., 2020). On the other hand, the author asserts that the integration of LLMs in the conventional RL paradigm, entailing the random exploration of the agent with its environment, will compensate for the occasional inaccuracies stemming from the language model's output (Carta et al., 2023; Brown et al., 2020). In essence, the author collectively asserts that this synergy bridges the gaps of each approach, rather than using these in isolation. To the best of the author's knowledge based on the conducted literature review of the related work in Section 2.4, this is a current gap in the literature.

6.2 Methodology

This section presents the methodology entailing the derivation and technical elaboration as well as the implementation of IGE-LLMs. Prior to the technical details of the IGE-LLMs framework, a short section outlining the technical preliminaries is presented. Thereafter, the core of the IGE-LLMs framework is detailed and how ROMAN's hierarchical architecture is extended for automated solving of intricate long-horizon robotic manipulation tasks via language-guided exploration.

6.2.1 Technical Preliminaries

First, technical preliminaries are outlined, which are considered important prior to the introduction and understanding of the proposed IGE-LLMs method. In particular, this section briefly presents related learning methods, including state-of-the-art (i) **count-based** and (ii) **prediction-based** intrinsic reward methods.

Markov Decision Process – The Markov Decision Process (MDP) is considered as a tuple (S, A, P, R, γ), whereby S and A represent the state and action spaces, respectively. $P(s'|s,a) = \Pr(s_{t+1} = s'|s_t = s, a_t = a)$ represents the transition function, specifying the probability of transitioning to the next state (s') given the current state (s) and the applied

action (a). The extrinsic reward function is given as $R(s,a,s')$, which provides the reward to the agent received after transitioning from state s to state s', given applied action a. The $\gamma \in [0,1]$ is the discount factor. The goal in this process is to learn a policy $\pi(a|s)$ that defines the probability of taking action a in state s such that the expected discounted returns are maximised as $\mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(s_t, a_t, s_{t+1}) \mid a_t \sim \pi(s_t)\right]$. To promote exploration in intricate tasks and settings, an additional function can be defined that complements the extrinsic reward function R with an intrinsic reward signal. This intrinsic reward should foster an exploratory edge, which should allow the agent to learn to explore interesting states.

Reinforcement Learning (PPO) – Similarly to ROMAN, denoting a policy π_θ as NN parameterised by weights θ, the PPO (Schulman et al., 2017) update at step k is given by: $\theta_{k+1} = \arg\max_\theta \mathbb{E}_{s,a\sim\pi_{\theta_k}} [L(s,a,\theta_k,\theta)]$, with a clipped loss function $L(s,a,\theta_k,\theta)$, with surrogate, value and entropy terms.

Count-Based Intrinsic Reward – Count-based intrinsic rewards encourage agents to visit states that have not frequently been encountered (Schäfer et al., 2022; Bellemare et al., 2016; Ostrovski et al., 2017). These rewards are inversely proportional to the visitation count of already encountered states. These are commonly represented as: $r_t^i = \frac{1}{\sqrt{N(s_t)}}$, whereby r_t^i denotes the intrinsic reward at time-step t and $N(s_t)$ the count of the encountered current state. In this process, higher intrinsic rewards are provided to the agent for those states that have been visited less frequently, thereby encouraging exploration.

Intrinsic curiosity module (ICM) – With ICM, an intrinsic reward signal (r^i) is introduced based on the ability of the agent to predict the consequence of its actions in a learned feature space. The intrinsic reward is formulated as $r_t^i = \alpha(\widehat{\phi}(s_{t+1}) - \phi(s_{t+1}))^2$. In this formulation, α represents a scaling factor, $\widehat{\phi}$ denotes the predicted representation of the next state given the current state and action, while ϕ denotes the actual representation of the next state. In this process, the model attempts to learn to encode information affected by the action performed by the agent. In essence, r^i represents the prediction error of the agent's estimate of the next state's feature representation. Consequently, the agent is encouraged to perform actions that maximise the error, exploring novel areas, consult Pathak et al. (2017).

Random network distillation (RND) – With RND a prediction-based intrinsic reward signal is introduced that is derived from the ability of the learned network, referred to as the predictor, to mimic a randomly initialised network which is referred to as the target network. The intrinsic reward r^i is formulated as $r_t^i = (\widehat{\phi}(s_{t+1}) - \phi(s_{t+1}))^2$. In this formulation, the terms $\widehat{\phi}$ and ϕ represent the predictor and target networks respectively for state s, consult Burda et al. (2019).

6.2.2 IGE-LLMs – Utilising LLMs as Intrinsic Assistance

As part of the last contribution of this thesis, a novel way of utilising LLMs to encourage efficient exploration in RL is proposed – IGE-LLMs. The IGE-LLMs framework, visualised in Figure 6.2, is tailored for RL tasks naturally challenged by long-horizons and sparse rewards, which necessitate an exploratory edge. In IGE-LLMs, given a state s_t of the environment at a given time t, the LLM is tasked with evaluating the potential future rewards of all actions $a \in A$ the agent is capable of performing. This evaluation is performed via a function $f : S \times A \rightarrow [0, 1]$, whereby S and A represent the state and action spaces respectively. The output of the action evaluation is represented as a continuous scalar value in the range of $\in [0, 1]$. For every pair of states and actions $((s_t, a))$, the LLM is tasked with assigning a rating $r^i = f(s_t, a)$. In essence, this rating represents the desirability of performing action a at state s_t. Hence, intuitively, f guides the agent towards actions that the LLM perceives as most beneficial given the current state s_t.

This process can be repeated practically once for every new state, prompting the LLM to rate the available actions the agent can perform. In this way, a dictionary (D) can be built and generated for every learning seed. By generating a dictionary as $D : S \times A \rightarrow [0, 1]$, computational costs of running inference on the LLM or Application Programming Interface (API)-related costs can be minimised and foster simultaneously efficiency. In this process, whenever the agent encounters a state s_t, it directly retrieves the corresponding intrinsic rewards for all actions from the dictionary D. In particular, if the state-action pair (s_t, a) already exists in D, which in essence implies this was previously encountered, the intrinsic reward is directly obtained as $r^i = D(s_t, a)$. Alternatively, if the state-action pair is not present in D, consequently meaning that this state was not yet encountered, the LLM is prompted to compute the intrinsic reward $r^i = f(s_t, a)$, which is subsequently stored in the dictionary as $D(s_t, a) = r^i$. In this way, the dictionary serves as an efficient guide for future actions of the agent in similar states by comparing whether the current

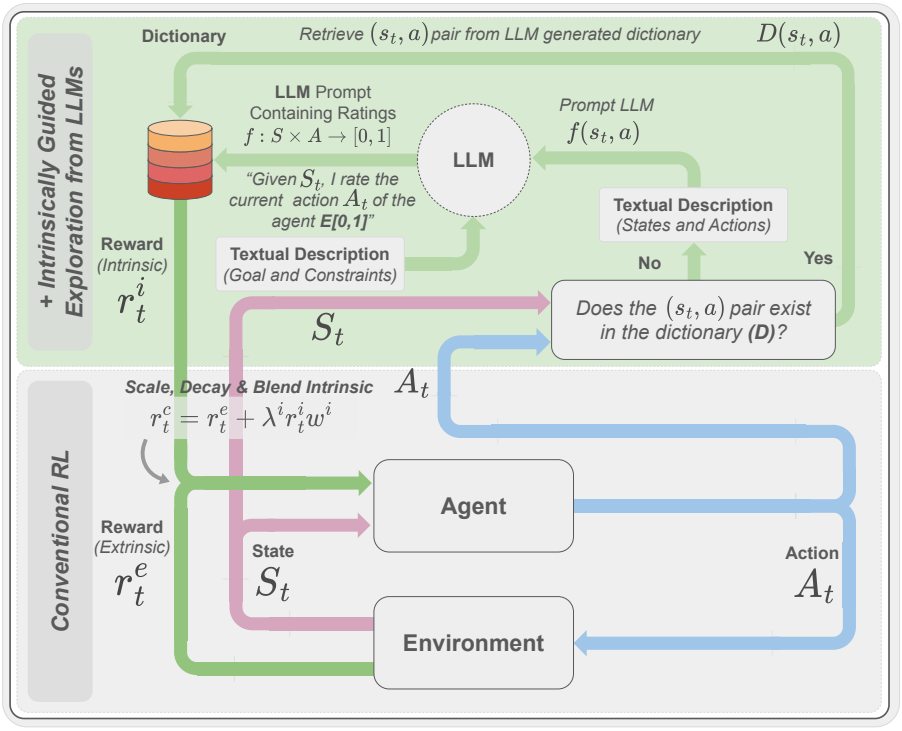

Figure 6.2: **Schematic illustrating the principles of IGE-LLMs and in particular the general and simplified overview of the framework.**

state-action pairs (s_t, a) were previously encountered.

By leveraging the existing RL paradigm, the total combined reward is given as $r^c = r^e + \lambda^i \cdot r^i \cdot w^i$. In this, the combined reward is represented as r^c, entailing the sum of the extrinsic (r^e) and intrinsic (r^i) terms. The intrinsic term is furthermore controlled by an intrinsic scaling factor denoted as λ^i, as well as a linearly decaying weight w^i. These two terms serve to guide the policy, without overfitting it to the intrinsic reward, due to naturally occurring inaccuracies stemming from LLMs (Brown et al., 2020). In this process, the agent receives a decaying over time dense intrinsic reward and gradually learns to converge towards the sparse extrinsic from the environment. The author hypothesises that this approach will encourage exploration in sparse reward environments and will further counterbalance the occasional inaccuracies from LLMs as these are merely utilised as an assistive intrinsic source and not as main policy drivers. Figure 6.2 visually depicts the general and simplified overview of IGE-LLMs.

As with ROMAN's HHL (see Chapter 5), for IGE-LLMs, PPO is chosen as the main RL algorithm. Despite being on-policy and generally less sample efficient than the off-policy SAC, it is preferred due to being less prone to instabilities and typically requires less

hyperparameter tuning than SAC (Schulman et al., 2017; Gu et al., 2017).

6.2.3 Extending ROMAN with IGE-LLMs

As part of this chapter, the ROMAN architecture (detailed in Chapter 5) is extended with IGE-LLMs, visualised in Figure 6.1. In particular, ROMAN's environment, which entails an intricate robotic manipulation task challenged by sparse rewards and long-horizons, presented itself as a promising basis to evaluate IGE-LLMs and related intrinsically-motivated methods.

The author of this thesis hypothesises that LLMs can be of notable benefit in hierarchical architectures, especially those tasked with completing notably intricate long-horizon robotic manipulation tasks. In such settings, correct sequential coordination and activation of a plethora of different macro-actions are necessitated to accomplish a long-horizon end-goal as seen in the works of Triantafyllidis et al. (2023b). Aggravating such settings further is that the completion of a sequential end goal is contingent upon the successful completion of other sub-tasks; for instance, one cannot retrieve an item from a drawer without first opening it. In the environment of ROMAN, there is a multi-sub-task-dependence, which in essence requires that every sub-task that is contingent on one another to be completed successfully. The same laboratory environment that ROMAN was evaluated upon is used, visually depicted in Figure 6.1 and Figure 6.3.A. This also includes the sequential task, detailed in Section 5.2.5.

Consequently, in order for the author of this thesis to test their hypothesis, the ROMAN architecture is extended, retaining all of its distinct specialising experts which are commonly seen in real-life (Billard and Kragic, 2019), treated hereinafter as **macro-actions**.

The author asserts that given the integration of IGE-LLMs on ROMAN's architecture and more specifically at its highest level – the gating network's (the MN) – the language guidance stemming from the framework will automate the supervision and orchestration of ROMAN's macro actions, offloading the necessity of human supervision at the hierarchy's highest level. In this way, ROMAN can retain its biologically-inspired intelligence for lower-level tasks (experts/macro-actions), yet the orchestration and supervision of those skills can potentially be automated by a capable language-guided framework fostering exploration. IGE-LLMs addresses this gap.

Implementation Details

As aforementioned, ROMAN's experts are treated henceforth as macro-actions. For more details regarding ROMAN's experts consult Section 5.2.6 and in particular Table 5.2 and Table 5.3. Macro-actions are maintained, allowing all methodologies compared in this chapter to be studied hereinafter on the gating network's ability (the MN), to coordinate these to solve intricate long-horizon sequential robotic manipulation tasks. It is asserted that utilising IGE-LLMs on ROMAN's gating network will leverage the LLM's useful context-aware reasoning for enhanced exploration. Since IGE-LLMs furthermore utilises the LLM as a mere assistive intrinsic signal alongside the extrinsic reward as per the RL paradigm, the author further hypothesises that this will counter-balance natural occurring inaccuracies of the language model's output. While Figure 6.2 visually depicts the general overview of IGE-LLMs, Figure 6.1 visually depicts the specific schematic of IGE-LLMs applied and extending the existing ROMAN hierarchy, consult Figure 5.2 for the architecture's overview.

Hybrid Learning and Characteristics of the Macro-Actions

ROMAN's NNs, treated as macro-actions, were trained in a hybrid manner as outlined in detail in Section 5.2.2. As a brief recap, the hybrid training combined BC (Torabi et al., 2018), RL (PPO) (Schulman et al., 2017), andGAIL (Ho and Ermon, 2016). Initially, the policies were warm-started with BC, thereafter updated using RL (PPO) with extrinsic (r^e) and intrinsic (r^i) rewards stemming from the environment and the GAIL discriminator respectively. Each NN received $N = 20$ demonstrations from keybindings, corresponding to velocity controlling the end-effector and its binary gripper state.

The proprioceptive states were identical for all NNs. Exteroceptive states differed, dependent on each NN's specialising skill. This, in turn, allowed each NN to focus on its own core exteroceptive information relevant to its sub-task, which is influenced from a neuro-scientific perspective whereby humans determine information relevance during motor tasks (Wolpert et al., 2011). All incorporated NNs in ROMAN's architecture shared identical actions, controlling the velocity of the robotic end-effector and the binary gripper state. A sparse, terminal reward (+1) was given upon completion of each network's specific sub-task goal. For more details consult Table 5.3.

6.3 Evaluation

IGE-LLMs and other state-of-the-art intrinsically-motivated methods, including LLM implementation methods were evaluated in two environments. First, a preliminary toy setting was utilised, targeting the challenge of RL exploration and thereafter a main environment entailing a complex sequential robotic manipulation task addressing both exploration and long-horizon challenges (ROMAN). The two environments utilised for the evaluation are visually shown in Figure 6.3. Each training seed utilising the LLM, corresponded to a different generated dictionary. This ensured randomness and diverse prompts from the LLM side. Lastly, a function $\phi(s)$ is employed to convert state vectors into textual descriptions for LLM input. This function was implemented given that direct state vectors might be uninterpretable for LLMs (Yu et al., 2023) and hence allow a high-level textual description of those states. To retain consistency and fairness with other baselines, any high-level description is also integrated and extended as binary values in the state vector. The prompts utilised for both environments can be found in the Appendices, in particular in Section B.1.

6.3.1 Illustrating the idea on a Toy Environment – DeepSea

Prior to proceeding to the main robotic task, a toy environment consisting of NxN grid-based environment is utilised – DeepSea. DeepSea is proposed as part of the Behaviour Suite (Bsuite) for RL whereby the challenge of exploration is targeted (Osband et al., 2020). In this environment, the agent starts at the top-left and aims to reach the goal located in the bottom-right grid. As part of the evaluation, increased grid sizes are evaluated ($N \in \{8, 64, 128, 192\}$), thus notably increasing the dimensionality and complexity of the environment. This environment is visually depicted in Figure 6.3.B.

States, Actions and Rewards – In regards to the state space (s_t), the agent position $(x1, y1)$ and target goal $(x2, y2)$ are given. The agent's actions (a_t) correspond to either moving down $(x, y+1)$ or diagonally down-right $(x+1, y+1)$. Reaching the goal yielded a sparse +1 terminal reward, while other bottom tiles incurred a -1 penalty.

LLM Implementation – The LLM is provided a prompt at each time-step of the DeepSea environment, corresponding to the same frequency of the agent. Similarly to the NNs having access to the positions of the agent and the goal, the LLM is provided identical

Figure 6.3: **The depiction of the two environments used for evaluating all methods in the chapter including the proposed IGE-LLMs framework. Figure A:** ROMAN, the main environment studying intricate robotic manipulation tasks entailing the correct sequential orchestration of a plethora of macro-actions. It is worthwhile to point out that all methods studied in this chapter were applied to ROMAN's gating network (MN) directly, with its experts in the hierarchy treated as macro-actions. Consequently, the evaluation here entails how well the MN is capable of inferring and orchestrating macro-actions in a robotic manipulation task challenged by both exploratory and long-horizon difficulties. ROMAN is challenged by exploration and long-horizons. **Figure B:** DeepSea, the preliminary grid-based environment. From left to right, increasing grid sizes of the DeepSea environment, significantly increasing the dimensionality and by extent the exploration necessitated to reach the goal. Blue, green and red items represent the agent, goal and traps respectively. **B.1-B.4** represent grid sizes *8x8*, *64x64*, *128x128* and *192x192* respectively. DeepSea is challenged by exploration. **Note on Visualisation:** For grid sizes *128x128* and *192x192*, agent virtual mesh size (in blue) is increased four-fold for enhanced visualisation purposes.

information in the form of textual descriptions corresponding to grid positions. The LLM is tasked to rate $E(0,1)$ the two possible actions (a_t) of the agent, given the current position of the agent and the goal's constant position located at the bottom right of the grid. Consult the Appendices (Section B.1) for the specific textual prompt descriptions provided to the LLM.

6.3.2 LLMs for Long-Horizon Robotic Manipulation Tasks

ROMAN's main environment is used as the evaluated task, entailing a set of intricate robotic manipulation tasks challenged by exploration and long-horizons, commonly encountered in robotics and physics-based interactions (Triantafyllidis et al., 2023b; Billard and Kragic, 2019). ROMAN's environment is depicted in Figure 6.3.A.

States, Actions and Rewards – In this environment, the gating network (the MN) was the primary NN used for the evaluation and its ability to orchestrate its macro-actions to achieve the long-horizon end goal. The end-goal, as per Chapter 5, consisted of the rack, to house the vial and both of these items to be placed on the conveyor, which had to be activated via switching on a button, for a visual overview consult Figure 5.1 and Figure 6.1. A softmax function as per ROMAN's architecture is implemented to its gating network as: $\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$. This allowed for the normalisation of the sum of weights of the seven macro-actions, ensuring these would not exceed union. Similarly, to ROMAN's original gating network, the state space of the extended version of ROMAN's gating network comprised the combined exteroceptive states of all its macro-actions in the hierarchy, allowing it to oversee the entirety of the environment. The proprioceptive states included the end-effector's position, velocity and gripper state. A sparse terminal +1 reward was awarded only upon completing the sequential end-goal. To allow a fair comparison with baselines, the gating network's state space is further extended in the form of binary values in the state vector, to accommodate the high-level textual description later on used for the LLM. Given the high complexity of the task, the author posits that the provision of dense intrinsic rewards stemming from the LLM with its context-aware reasoning will aid learning in intricate long-horizon manipulation tasks.

LLM Implementation – Similarly to DeepSea, the LLM is provided with a prompt at each time-step equivalent to the frequency of the gating NN (1Hz) in the main robotic environment. The LLM is given high-level textual descriptions which correspond to the

binary states of the primary OIs in the scene, which as aforementioned, is also integrated as binary values in the state vector to ensure fairness with baseline comparisons. At each time-step the LLM is tasked to rate $E(0,1)$ the seven possible macro actions (a_t), based on the textual description of the OIs, rating as such the state-action pairs (s_t, a). The description of those OIs is whether these are present or absent from the scene, based on their visual detection from the vision system detailed in Section 5.2.3. Consult the Appendices (Section B.1) for the specific textual prompt descriptions provided to the LLM.

6.4 Results and Value of IGE-LLMs

This section outlines the result of IGE-LLMs, conventional learning methods and existing intrinsic approaches. A comprehensive evaluation is performed to highlight the value of the proposed framework for a better understanding of its applicability and potential in addressing intricate learning tasks.

6.4.1 Evaluation Approach

The evaluation is performed across **five** distinct steps, which are outlined below.

Evaluation I: IGE-LLMs, Conventional Learning and Intrinsic Methods – Initially, a comparison is made amongst conventional RL utilising extrinsic rewards (**RL**), thereafter combining it with intrinsic methods, including prediction-based (**ICM**, **RND**), **count-based**, as well as the proposed **IGE-LLMs** of this work. This will allow for a thorough understanding of how these learning-based methods cope against the two studied environments; one challenged by exploration and the other by both exploration and long-horizon challenges.

Evaluation II: The Modularity of IGE-LLMs – Next, the proposed **IGE-LLMs** method is combined with existing prediction (**IGE-LLMs wICM**) and count-based (**IGE-LLMs wCount**) intrinsic methods. In this way, the proposed method's modularity can be demonstrated and how it can be combined and complement existing intrinsically motivated methods.

Evaluation III: LLMs as Main Policy Drivers – Furthermore, to underscore the inherent limitations of occasional prompt inaccuracies stemming from LLMs (Carta et al.,

2023; Brown et al., 2020), the direct outputs stemming from the LLM are compared, without their inclusion in any learning or training based setting. To this end, the state-of-the-art reasoning referred to as Chain of Thought Reasoning (CoT) from the works of Wei et al. (2023) is utilised. CoT is incorporated due to its ability to elicit notably improved arithmetic, commonsense, and symbolic reasoning. For both direct **LLM** and **LLM with CoT** outputs, the **argmax** is used to directly retrieve the highest rated output from the LLM's prompts, without using it as intrinsic guidance as with **IGE-LLMs**. In this way, the capability of the LLMs can be tested directly, as main policy drivers rather as assistance. To further incorporate some randomness into the decision-making process with the argmax, when identically rated actions by the LLM are present in the dictionary, a random selection is made amongst those occurrences to ensure diversity in the selection process of those ratings.

Evaluation IV: Sensitivity Analysis of IGE-LLMs – Since the IGE-LLMs framework is at its core an intrinsic-motivated method utilising LLMs as an intrinsic reward source, it is important to validate its sensitivity. In particular, a common challenge when employing intrinsic-motivated exploration is the sensitivity to the intrinsic scale λ, consult Schäfer et al. (2022). Consequently, a sensitivity analysis of IGE-LLMs is performed, by evaluating a varying number of logarithmic spaced intrinsic scaling parameters (λ). This in turn will provide a comprehensive analysis of how robust the framework is to such scaling parameters.

Evaluation V: Robustness Evaluation – Lastly, a robustness test of all trained models in inference is performed over increasing uncertainty levels and increased horizons to comprehensively study their robustness. This is performed on ROMAN's environment, providing in turn an extensive overview of how well the studied methods cope with increased complexities, naturally observed in real-life robotic tasks (Davchev et al., 2022; Triantafyllidis et al., 2023b).

6.4.2 Results

The results are listed across **five** distinct steps, which are outlined below.

Results I: Utilising LLMs as Intrinsic Assistance – The results for both DeepSea and the main ROMAN environments are depicted in Figure 6.4. From the environment returns, it is initially inferred that existing intrinsic reward methods (r^i) are overall beneficial in promoting exploration in sparse environments compared to just the use of

conventional **RL** entailing only extrinsic rewards (r^e). Nevertheless, it is still observed that most intrinsic methods still struggle with increased dimensionality and horizons. Most notably, it is shown that **RLwCount** performs well for DeepSea, yet shows significantly lower returns in increased dimensionality ($192x192$). In ROMAN's environment, **RLwCount** attains less than approximately half of the normalised return. **RLwRND** and **RLwICM** both struggle in DeepSea for grids beyond and inclusive of $64x64$. In ROMAN's environment, **RLwRND** exhibits notable drops in returns, while **RLwICM** notably higher returns, still not effectively close to the maximum attainable. In contrast, IGE-LLMs consistently performs well in DeepSea across all tested grid sizes regardless of complexity. Most importantly, IGE-LLMs outperforms by a notable margin all other methods in ROMAN's environment. These findings suggest that utilising LLMs as intrinsic assistance appears to be promising, especially in environments entailing sparse rewards and long-horizons, as commonly observed in intricate robotic manipulation tasks.

Results II: Combining IGE-LLMs with Existing Intrinsic Methods – Part of the next evaluation, was to study the modularity of IGE-LLMs and in particular how well it can be combined with existing intrinsic based reward methods. From Figure 6.4, results show that IGE-LLMs is capable of being combined with existing prediction (**IGE-LLMswICM**) and count-based (**IGE-LLMswCount**) intrinsic rewards. In particular, it is shown that using both **RLwICM** and **RLwCount** in isolation renders low returns with increasing grid sizes in DeepSea and in particular in ROMAN's intricate environment. Contrarily, when these are coupled with the proposed IGE-LLMs framework, these perform notably better than before. These results suggest that IGE-LLMs can also be combined and complement existing prediction and count-based intrinsic methods, subsequently highlighting its modularity and robustness.

Results III: Constraints of Direct use of LLMs in the Decision Making – Part of the main hypothesis of this work was that utilising LLMs in IGE-LLMs as an assistive intrinsic proxy reward, alongside the traditional extrinsic in the conventional RL setting, will compensate for the LLM's occasional and naturally occurring inaccuracies (Carta et al., 2023; Brown et al., 2020). This hypothesis stemmed from the fact that IGE-LLMs's framework will not cause the policy to learn sub-optimally as the replies stemming from the language model are simply used for guidance, with the extrinsic reward still eventually being the main policy driver. To thoroughly test this hypothesis, the direct integration of LLMs into the decision-making process is studied as part of the

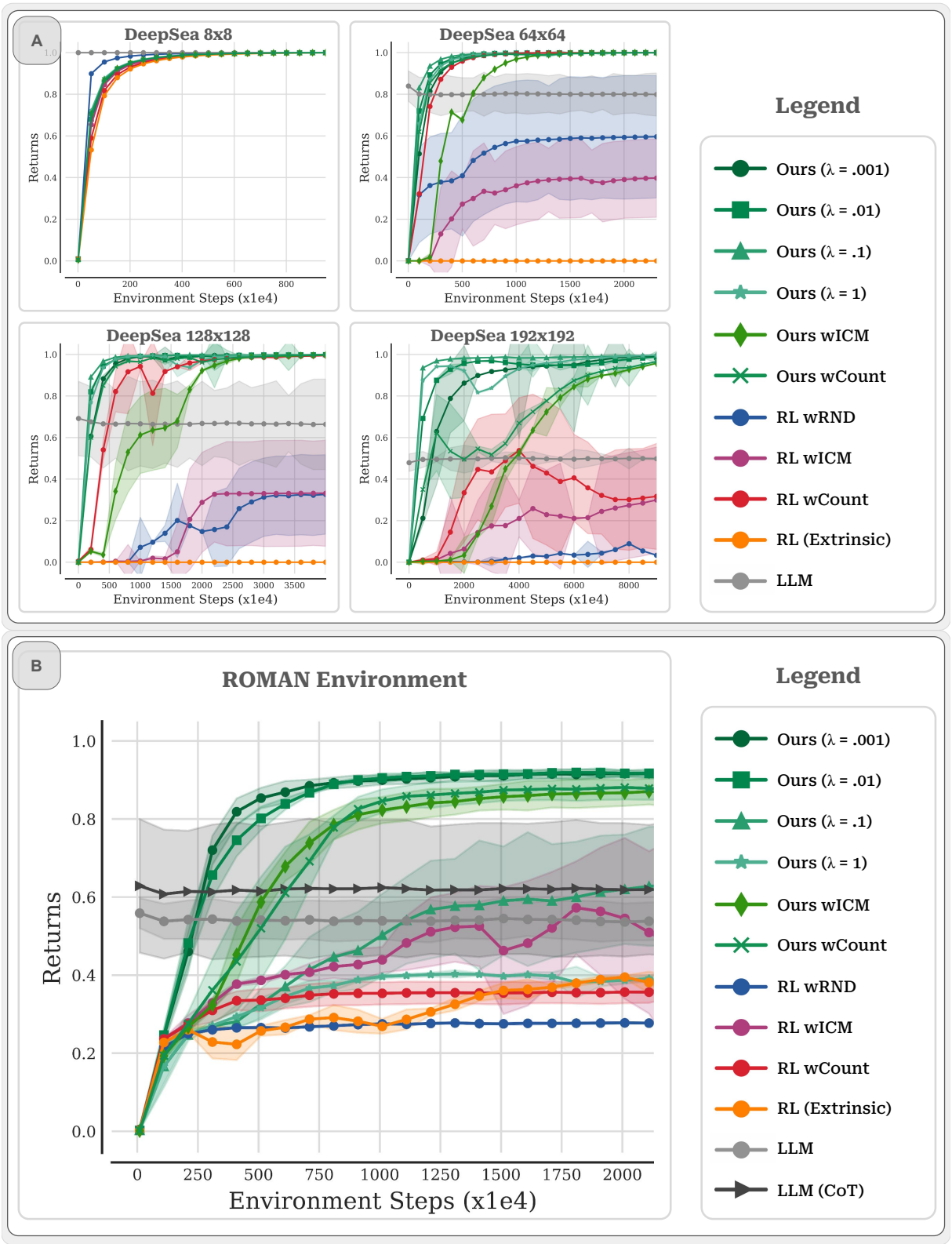

Figure 6.4: **The normalised evaluation returns for the studied preliminary and main robotic environments.** The shaded region depicts the standard deviation (σ) around the mean, indicating the variability of rewards across different seeds. Smoothing and averaging are applied for enhanced visualisation purposes for all plots. **Figure A:** The normalised returns for DeepSea for varying grid sizes NxN. Results stem from $n = 5$ seeds for *8x8* and *64x64* and $n = 3$ seeds for *128x128* and *192x192*. **Figure B:** The normalised evaluation returns for ROMAN. Results stem from $n = 5$ seeds. **Note:** The LLM with the incorporation of CoT (Wei et al., 2023), is applied in the environment of ROMAN.

next evaluation. From Figure 6.4, it is inferred that unlike IGE-LLMs, direct reliance on LLMs without their inclusion in a learning setting is discouraged. In particular, results show that even advanced LLMs (with *GPT-4* being the model of choice), are prone to occasional inaccuracies, exhibiting, in turn, significant errors. The argmax results of direct **LLM** usage in both the DeepSea and ROMAN environments show that these are inadequate to account both against increased dimensionality, exploratory necessity and long-horizon tasks. Furthermore, even when incorporating CoT as with the work of Wei et al. (2023) in the LLM (see Figure 6.4.B), it is shown that (**LLMs (CoT)**), is still effectively unable to attain high returns in ROMAN's intricate robotic manipulation task. It is thus concluded that directly relying on LLMs should be avoided, which is in line with (Carta et al., 2023; Brown et al., 2020). Instead, results show that utilising LLMs as an assistive intrinsic signal is preferable. This is where the value IGE-LLMs is underlined. In this way, innate interactions of the agent with its environment, as per the RL paradigm, can counteract the inherent limitations when directly relying on and integrating LLMs in the decision-making process.

Results IV: Sensitivity Analysis of IGE-LLMs – From the related work, it is inferred that when employing intrinsic-motivated methods, a common challenge is the sensitivity to the intrinsic scaling parameter λ^i, consult Schäfer et al. (2022). As IGE-LLMs utilises at its core LLMs as an assistive intrinsic reward, it is deemed important to investigate the framework's robustness against varying values of λ^i. Hence, for sensitivity analysis, different logarithmic spaced scaling parameters $\lambda^i \in \{.001, .01, .1, 1\}$ are evaluated. In this way, the robustness of IGE-LLMs can be verified and tested, for intrinsic scaling parameters (λ^i) that would drive the intrinsic reward r^i to be from a fraction, equal to exceeding strengths of the extrinsic reward r^e. From Figure 6.4.A, it is inferred that IGE-LLMs is robust to all studied λ^i values in DeepSea, performing consistently well regardless of the scaling. Nevertheless, it is shown that while IGE-LLMs performs well for $\lambda^i \in \{0.001, 0.01\}$ in ROMAN's environment (refer to Figure 6.4.B), a notable drop in returns is observed with those equaling and exceeding the r^e ($\lambda^i \in \{0.1, 1\}$). The drop in returns based on high λ^i values can be attributed to the intricate nature of the robotic task studied in the ROMAN environment as well as the overemphasis on the LLM's outputs. The latter is due to high λ^i values eventually and inevitably overemphasising the r^i term, possibly overshadowing the extrinsic term r^e. From the findings on **Results III**, it is indeed shown that an overemphasis on the LLM's output is discouraged due to its occasional inaccuracies (Carta et al., 2023; Brown et al., 2020), as evidenced by

the argmax of the LLM and LLMwCoT results. Nevertheless, from the results at hand, IGE-LLMs still maintains fair amounts of insensitivity to large r^i values.

Results V: Inference Results - Robustness Test Against Increased Uncertainty and Horizons – As part of the final evaluation, all tested models are run in inference and evaluated for robustness in the ROMAN environment. In particular, similarly to Chapter 5, Gaussian distributed noise is added to all exteroceptive states in ROMAN's gating network (the MN), corresponding to positional observations. Uncertainty levels from 0cm to up to $\sigma = \pm 2$cm, in 0.5cm increments are studied. With this evaluation, both (i) increasing uncertainty levels and (ii) horizons are studied. Results are shown in Figure 6.5. From the results, it is initially observed that utilising intrinsic rewards (r^i) complements RL only utilising extrinsic rewards (r^e). Upon further investigation, it is shown that **RL**, **RLwRND** and **RLwCount** appear to struggle with longer horizons, with the most notable result being Cases 4 to 7, which are effectively unattainable. On the other hand, **RLwICM** shows notably higher success rates both against uncertainty levels and horizons compared to **RL**, **RLwRND** and **RLwCount**, yet errors increase significantly for noise levels exceeding $\sigma = \pm 0.5$cm and especially for increased horizons. In contrast, IGE-LLMs, maintains high success rates across increasing uncertainty levels and horizons, outperforming by a significant margin RL and other intrinsic methods.

6.5 Discussion on Utilising LLMs for Exploratory Potential and Emergent Behaviour

This chapter presented the final contribution of this thesis, the IGE-LLMss framework. Addressing the gap in the current literature, IGE-LLMs is a novel approach utilising LLMs as an assistive intrinsic source, fostering and promoting exploration in RL tasks challenged by sparse rewards and long-horizons. Validated on a preliminary and subsequently on ROMAN's intricate long-horizon robotic manipulation environment, the results of this chapter showed that the IGE-LLMs framework outperformed existing intrinsic methods and underlined and overcame the shortcomings of the direct use of LLMs in the decision-making process, even when employing state-of-the-art CoT reasoning as with the works of Wei et al. (2023). Exhibiting consistently higher environment returns over existing methods, IGE-LLMs harnesses the textual insights of LLMs as the means of guidance and the random interaction of the agent with its environment via the conventional RL

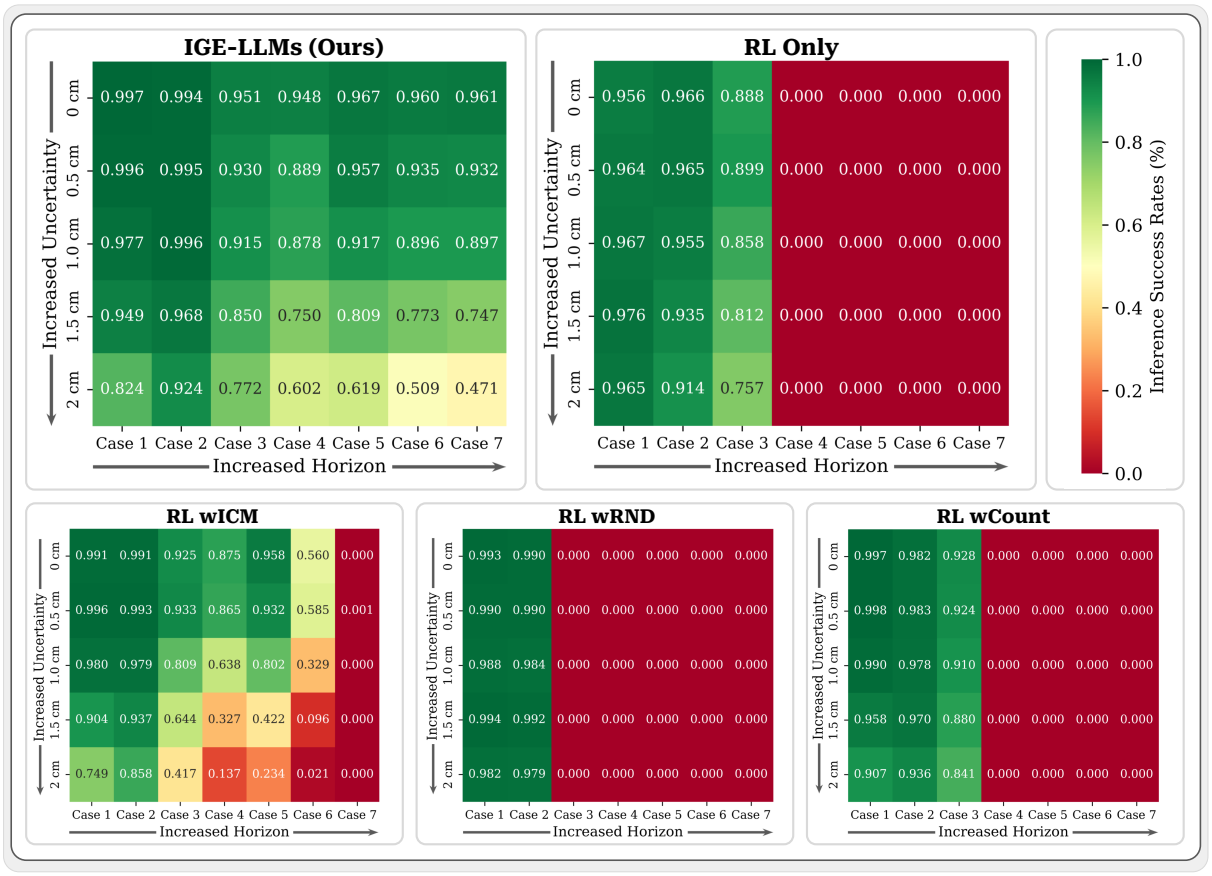

Figure 6.5: **Robustness test of all studied models for the ROMAN environment including the IGE-LLMs framework.** The figure depicts the inference results of the gating network's (MN), ability to orchestrate all entailed macro-actions in the hierarchy of ROMAN, against increased uncertainty levels and increased horizons. Each heat map visualised represents the distinct models studied. In particular, five distinct models are presented, namely IGE-LLMs, RL, RLwICM, RLwRND and RLwCount. Within each heat map, the x-axis represents the task horizon, ascending from left to right, while the y-axis represents exteroceptive noise levels, ascending from top to bottom. Each cell stems from 1,000 episodes.

paradigm, complementing each approach and overcoming the limitations of each when otherwise used in isolation. Moreover, IGE-LLMs's modularity was underlined by its ability to be integrated and complement with existing intrinsic reward methods, its insensitivity to most scaling parameters underscored and its sustained robustness highlighted against increased uncertainty levels and horizons over other methods.

It was furthermore shown that IGE-LLMs framework is highly modular, reflected by the framework's ability to be implemented directly on ROMAN's hierarchical architecture in a straightforward manner. In particular, this chapter illustrated how ROMAN can be modified such that the main gating network (the MN), can be substituted by IGE-LLMs, while retaining its specialising and distinct experts, treated thereafter as macro-actions. Effectively, this modification allowed the ROMAN architecture to still maintain its biologically inspired embodied intelligence perspective via its experts which were trained in a hybrid manner (as outlined Figure 6.1), yet the higher-level orchestration of those was automated by IGE-LLMs (general overview seen Figure 6.2), providing invaluable context and common sense aware reasoning. Ultimately, results showed that ROMAN with IGE-LLMs renders an autonomous process capable of planning and addressing intricate long-horizon tasks efficiently and robustly. Furthermore, given that a common challenge is that state vectors and lower-level actions may be uninterpretable for LLMs (Yu et al., 2023), retaining the macro-actions of ROMAN and instead utilising LLMs for higher-level orchestration of these via the proposed IGE-LLMs framework at the gating network's level, presented itself as a suitable mitigation strategy to this challenge.

Overall, from the findings and results of this chapter, it is inferred that IGE-LLMs presents itself as a promising framework for addressing the exploration and long-horizon challenges prominently encountered in complex robotic manipulation tasks. In the next and final chapter of this thesis (Chapter 7), the limitations, future work and areas of improvement for IGE-LLMs and more broadly the utilisation of LLMs are detailed.

Chapter 7

Conclusion

The vision of human-robot symbiosis, wherein both entities augment each other's capabilities, has been a long-standing aspiration. Combining the reasoning and more broadly the intricate and highly advanced sensory-motor system of humans, together with the augmented physical strength, durability, and precise repetitive task executions of embodied intelligence is a grand research endeavour at the intersection of human ergonomics, robotics, and AI. Bridging these capabilities from a biological and artificial perspective remains an ideal vision for future human-machine collaboration. Such bio-digital synergy holds the promise of an era where humans and machines can harmoniously coexist.

To inch closer to the aforementioned endeavour, this thesis presented advancements for the intricate understanding of both the sensory-motor system of humans and with those insights the development of biologically-inspired embodied intelligence was accomplished. While this thesis presented a first blueprint towards this endeavour, more research is necessitated to align the remarkable capabilities exhibited by humans closer to machines. This will subsequently enable the creation of embodied intelligence that mirrors, or perhaps even surpasses, the remarkable capabilities of humans.

This chapter encapsulates the strides taken in this doctoral thesis towards that visionary horizon by delving into the intricate human sensory-motor system and developing embodied intelligence rooted and inspired by biology. The central aim of this doctoral thesis was two-fold. First, the aim was to understand, harness and quantify the human sensory-motor system reflected in intricate motor tasks commonly seen in 3D interactions entailing MR technologies and teleoperation tasks. Secondly, inspired by the first aim,

the second was to develop appropriately bio-inspired embodied intelligence for hierarchical task learning (ROMAN), furthermore aided by language (IGE-LLMs), to address the intricate domain of long-horizon, sequential of sparse rewards manipulation tasks in robotics and more broadly in machine learning.

In the remainder of this chapter, Section 7.1 delves into a succinct summary of the four pivotal contributions across Chapter 3 to Chapter 6. Following that, Section 7.2 acknowledges the research's limitations and sets the stage for prospective research explorations necessitated to further narrow the gap of the central aims of this doctoral work. Both subsequent sections provide articulated summaries and outline future work, separated into distinct sections for clarity purposes.

7.1 Summary

In summary, this research has presented advancements towards the human sensory-motor system, machine learning, and robotics. Key contributions span from the multifaceted exploration of the intricate human sensory-motor system, leading to the inception of biologically-inspired embodied intelligence, moreover solidified and enhanced by language-guided frameworks. These novel contributions have provided a first step towards bridging the gap between human and machine capabilities in intricate tasks, while further solidifying the capabilities of AI motivated throughout biology.

In particular, in Chapter 3, the human sensory-motor system was investigated and how it can harnessed and elucidated. Through a comprehensive series of experiments assessing a plethora of sensory modalities, a correlation was established as to which sensory states contribute towards effective motor actions. Next, in Chapter 4 the human sensory-motor system was further evaluated, with the primary focus of deriving a metric that can quantify and measure such motor movements and actions stemming from humans. By exploring a myriad of existing evaluation tools and formulations, this research derived a novel formulation that can capture the intricate motor movements and actions exhibited by humans in higher-dimensional space of varying DoF. The penultimate and larger part of this thesis was presented in Chapter 5. Motivated by the previous two contributions delving into the human sensory-sensory system, Chapter 5 derived a biologically-inspired hierarchical embodied intelligence architecture – ROMAN. ROMAN's value was underlined by its ability to solve notably intricate long-horizon sequential robotic manipulation

tasks, by virtue of its employed HHL approach, rooted in biology, as well as the employed task-decomposition and overall architecture characteristics. Finally, in Chapter 6, motivated throughout from a biological perspective, the final aim was to harness the power of language. To this end, the IGE-LLMs framework was presented, to further solidify embodied intelligence and bridge the gap closer to the capabilities exhibited by humans. By leveraging LLMs as an assistive reward, IGE-LLMs exhibited notably improved performance over related learning methods, harnessing the context and common-sense aware reasoning from textual insights, to promote an efficient exploration strategy. The four aforementioned key contributions are outlined in distinct paragraphs below.

Chapter 3 delved into the human sensory-motor system with the hopes of understanding and establishing a correlation between sensory stimulation and effective motor actions. To this end, a novel and comprehensive full-factorial comparison of the most widely used sensory cues of vision, audition and somatosensory feedback was conducted to establish patterns between sensory stimulation and motor performance of varying complexity tasks. To the best of the author's knowledge, this was a first-of-its-kind comparison of the most widely used sensory modalities of vision (monocular display vs stereoscopic VRHMD), audition (presence vs absence) and somatosensory (presence vs absence) feedback. The contribution of these modalities towards human motor performance was evaluated across a variety of different metrics reflected upon tasks of varying amounts of spatial difficulty. From the results and findings, it was inferred that motor task performance was overall increased across all measurements by 40% when utilising stereoscopic over monocular vision, with the presence of somatosensory feedback contributing to an additional 10% over all incorporated metrics, while auditory feedback showed a marginal improvement of 5% limited to spatial accuracy. Moreover from these results, it was also concluded that instead of solely depending on a single modality, the addition of more sensory modalities, i.e. multimodality, with either bi-modal or even more so tri-modal feedback contributes to a positive increase in functional human motor actions and performance. Additional findings revealed that uni-/bi-/tri-modal feedback states and their variations, correlate to varying motor action performance under increasing spatial complexity manipulation tasks. As a result of this research, future researchers and designers should identify and prioritise certain modalities and their different combinations in order to design effective and efficient multimodal interfaces to effectively harness the human sensory-motor system for improved motor task performance.

Chapter 4 delved similarly to the human sensory-motor system, with an emphasis on measuring and quantifying meaningful motor actions and motor movements exhibited by humans. In particular, the derivation of a novel human predictive model was presented based on Fitts' original formulation. This novel metric was extended to full 3D space entailing intricate DoFs, including combined translational and rotational movements. Moreover, as per the original paradigm of Fitts' original law, it is supported that the derived metric is capable of capturing pointing types of motor tasks, with the added advantage of also capturing physical-based interactions as seen in real life and robotic manipulation. From the results across four in total increasing spatial complexity motor tasks, it is supported by the data that the derived formulation outperformed all known model extensions to date. This held especially true amongst the most complex spatial settings as those encountered in full 3D space. More specifically, it was observed that in the most basic form of 3D translational pointing and manipulation (*Motor Task 1*), existing model extensions performed adequately well. Thereafter, these were mostly unable to capture the addition of spatial arrangements of directions and inclinations (*Motor Task 2*). Subsequently, most model extensions showed further decreased model fittings during isolated rotational motions (*Motor Task 3*). Finally, all models to date exhibited overall very poor fitting results during combined translational and rotational with spatial arrangements of directional and inclinational motions (*Motor Task 4*). The derived metric, as shown, is capable of predicting the required time to complete a plethora of complex higher DoFs 3D movements based on the intricately associated spatial variables entailed in higher dimensional space. Consequently, the formulation is capable of combining time and spatial-based metrics under a single model. This in turn, as inferred by the related work of Triantafyllidis and Li (2021b), increases overall inter-study comparability and due to its advantage of being based on Fitts' original law, generalisation towards quantifying human motor actions and performance is to a great extent increased. Overall, from the results it is supported that this metric can capture the intricate human motor movements and associated spatial variables as seen during robotic teleoperation, 3D MR-based interactions (notably, 3D user interfaces) and more broadly human ergonomics.

Chapter 5, the penultimate and larger part of this thesis presented the proposal of a biologically-inspired learning architecture – ROMAN. ROMAN at its core, utilised a HHL architecture capable of solving notably intricate long horizon sequential manipulation tasks. Within ROMAN's framework, the internal hierarchical model and the plethora of entailed experts, treated as macro actions, are governed by a central gating network

which is referred to as the MN. The ability of the MN to orchestrate the sequential activation of the entailed specialised manipulation experts to generate their correct activation is attributed to the internal hybrid learning architecture as well as the higher-level task decomposition of these experts. These employed strategies resulted in the offloading of low-level task supervision for MN, which allowed, in turn, the gating network to focus instead on primary higher-level scene understandings to accomplish the sequential end goal. By fusing a plethora of different learning paradigms, the resultant framework achieves a balance between exploitation and exploration, rendering the internal policy to achieve high success rates even amongst the most intricate sequential settings. Most importantly, the exhibited versatility, as well as dynamic adaptability during out-of-distribution states, underlined the value of ROMAN and its capabilities for autonomous failure recovery. Ultimately, this adaptability of ROMAN rendered the hierarchy of particular value in settings necessitating long-horizon sequential actions with adaptive motor skills, in which conventional state machines would face notable challenges. From the results, it is inferred that the hierarchical formulation of ROMAN and its internal HHL architecture resulted in (i) robustness against the presence of notably intricate long horizon sequential tasks, (ii) maintained high-performance increasingly high exteroceptive observational uncertainties, (iii) yielding consistently high success rates even on a vision-based detection module whereby ROMAN was not trained on, (iv) high performance even with fairly small number of demonstration data and finally (v) adaptability to cases beyond those encountered in the demonstrated sequence exhibiting recovery capabilities. These findings suggest the potential importance of ROMAN and its implications for intricate long-horizon sequential tasks, necessitating adaptive motor skills.

Chapter 6, the final part of this thesis, delved into the intricate dimension of LLMs to harness the power of language. By investigating and eliciting the textual reasoning of LLMs, the final contribution of this work further solidified the gaps between humans and machines, leading to the proposal of IGE-LLMs. IGE-LLMs is a novel framework leveraging LLMs as an assistive intrinsic source, tailored for promoting guidance and fostering exploratory potential in intricate RL and robotic tasks challenged by sparse rewards and long-horizons. Novel, intuitive and modular, IGE-LLMs fosters exploration by utilising the useful context and task-aware reasoning from LLMs as an assistive intrinsic proxy reward within the conventional RL setting. In this way, IGE-LLMs complements existing learning methods challenged by sparse rewards and long-horizons, while simultaneously overcoming the inherent limitations of LLMs, (Carta et al., 2023; Brown et al., 2020).

The latter was made possible by utilising LLMs as mere guidance and decaying the intrinsic signal by the LLM so that the sparse extrinsic reward would eventually become the main policy driver by leveraging the existing RL paradigm. IGE-LLMs was validated on two environments; one preliminary grid-based environment challenged by exploration and subsequently ROMAN's main environment entailing an intricate robotic manipulation task, challenged by both exploration and long-horizons. The results showed that IGE-LLMs exhibited notably higher performance than just the use of extrinsic rewards (RL) even when coupled with existing intrinsic reward methods. Moreover, results also revealed the shortcomings of the direct application of LLMs in the decision-making process, highlighting as such the value of using these as guidance (as with IGE-LLMs) rather than directly as main policy drivers. Moreover, IGE-LLMs's modularity was underscored by the ability of the framework to be combined and complement existing intrinsic methods, while also extending and enhancing the performance of complex hierarchical architectures (ROMAN's MN). Furthermore, IGE-LLMs's insensitivity to most intrinsic scaling parameters was featured by evaluating it against increasingly spaced logarithmic scales. Lastly, by running all models including IGE-LLMs's in inference, a robustness test revealed that the proposed framework performed consistently well and retained high success rates both against increased uncertainties and horizons. Collectively, these findings underline the value of IGE-LLMs for tackling the exploration and long-horizon challenges, inherent in ROMAN's environment and more broadly in intricate robotic manipulation tasks.

7.2 Limitations and Future Work

The contents and research outcomes of this thesis have provided novel findings on (i) understanding and harnessing the human sensory-motor system via a comprehensive analysis of multisensory states and their effect on motor actions, (ii) the quantification of the human motor actions and movements reflected on increasingly intricate spatial motor tasks, to (iii) inspire the development of a biologically-inspired HHL architecture – ROMAN – to address intricate long-horizon sequential manipulation tasks, that can lastly be (iv) enhanced with the power of language-guidance, leading to the proposal of IGE-LLMs. Nevertheless, improvements to these four key aspects were identified and were left part of this thesis's future work. With the aim of further narrowing and solidifying the gap between the capabilities exhibited by humans and the computational power of AI, further research is necessitated to understand the intricate human sensory-motor system and the

implications for the development of efficient biologically-inspired embodied intelligence. Inspiration from biology has the potential to endow embodied intelligence with the capabilities necessitated to meet intricate environmental demands seamlessly.

As a result of this research thesis, certain important limitations and improvements were identified. These are outlined in this particular section, separated into individual paragraphs corresponding to the distinct Chapter 3 to Chapter 6 for clarity purposes.

Chapter 3 presented a thorough overview of how the human sensory system can be understood and harnessed via appropriate multi-sensory stimulation, with certain aspects identified for future improvement. As part of future work, multisensory synchronisation during multimodal stimulation (whether bi-modal, tri-modal or more) is important. Most notably out-of-synchronisation sensory cues can reduce spatial and temporal immersion, which would to a great extent nullify the benefits of using multimodal stimulation according to Popescu et al. (2002); Richard et al. (1994). The implications of this are of vital importance due to the naturally occurring sensory delays and latency present in current hardware which may lead to some extent in the de-synchronisation of sensory feedback states. To which extent such latency during sensory stimulation can be measured and quantified is an important research direction to consider. Furthermore, the sensory design decision surrounding the specific sensory modalities is of equal importance during increased functional task complexity (Sigrist et al., 2013). For instance, whether visual feedback is provided in monocular or binocular viewing, auditory feedback in the form of concurrent or terminal feedback or somatosensory stimulation in the form of positional control, proximity or collisions, may all very much lead to different findings. Moreover, as part of the work undertaken in this research the most popular means of hardware devices to induce sensory stimulation were considered. Particularly, for the visual stimulation, two states were considered; monocular vs stereoscopic stemming from a generic display monitor and a VRHMD respectively. However, it would also be invaluable to further investigate purely monocular and stereoscopic stimulation on a single device and limit the analysis to just this perspective. This would also apply to other means of haptic stimulation e.g. force-feedback vs vibrotactile gloves and other possible permutations.

Chapter 4 presented comprehensive research regarding the human sensory system, focusing on the evaluation of the entailed spatial variables, to quantify high dimensional with 6DoFs motor movements and motor actions. This allowed in turn to derive a novel metric to measure and quantify the motor actions seen in the sensory-motor system in

humans. An emphasis was directed towards Fitts' law and the most popular extensions of the law, compared and evaluated across motor tasks of increasing spatial complexity. This allowed, in turn, to comprehensively analyse which spatial variables contributed towards motor performance, ultimately leading this research work to derive a novel metric extension that can model intricate human motor actions and movements. However, as with most work on extending Fitts' law, this work investigated seated, limited to upper-body human motor movements (Triantafyllidis et al., 2021). For future work, it would be invaluable to further evaluate the derived formulation on lower and upper body movements such as during locomotion entailing the simultaneous manipulation of objects. These biomechanical variations may lead to different findings yet more research is necessitated to confirm this. Due to the already extensive comparison of the spatial variables entailed in full 3D space investigated in this work, the research evaluation was limited to a specific range of those spatial variations. For future work, adding further levels of such spatial variations, for instance, further translational or rotational variables, would moreover aid in solidifying the derived metric. Additionally, adding more complex convex shapes in the experiments, e.g. non-primitive 3D objects such as a toy car would allow the investigation of more spatial variations, especially in regards to rotations compared to the cube object used in this work. The cube object was used to study the effectiveness of motor tasks with the most common and primitive 3D object possible, while also allowing the evaluation of a small range of rotational variations. Moreover, further research should also be redirected towards the different choices of distances that are readily available as well as the orientation adjustments and how these are integrated into combined movements during tasks like picking and placing. Lastly, other aspects that could influence any derived metric attempting to quantify motor movements could be for instance the different grasping types one could employ such as those seen in the SHAP (consultLight et al. (2002)), as well as the utilisation of the different robotic hand end-effector designs or the plethora of different input devices available. Nonetheless, this research was a novel attempt towards the derivation of a higher dimensional formulation with a basis on a well-established metric (Fitts' law), to comprehensively evaluate full-3D movements and motor actions, as commonly seen in MR-based interactions, robotic teleoperation and more broadly for settings necessitating carefully designed human ergonomics.

Chapter 5 presented the derivation of the novel biologically-inspired HHL ROMAN architecture tailored for intricate long-horizon tasks. As part of the future work, certain aspects were identified to further improve upon the derived ROMAN architecture. Foremost,

as it was inferred and identified from the conducted t-SNE, there were certain similarities in the exteroceptive vector of the MN. These similarities, in turn, yielded a slight inconsistency and in particular lower success rates for otherwise less intricate, smaller-horizon sequential tasks. A potential mitigation for this problem would be to endow ROMAN with some form of "memory mechanism", either explicitly by expanding the observation space to include a buffer of its past actions, or implicitly by employing some recurrent NN architecture for the MN. This was left as part of the future work of this thesis. Moreover, another potential future improvement would be the derivation of the lowest possible level of "fundamental" manipulation skills via a lower-level task decomposition as opposed to ROMAN's experts. While the derivation of higher-level experts in ROMAN allowed the MN to be offloaded by lower-level task supervision and instead focus on higher-level scene understandings, it may be potentially beneficial in terms of re-using demonstrations across a plethora of different expert NNs in a hierarchy (i.e. macro-actions), as supported by Fox et al. (2019). Moreover, automating the decomposition of tasks via LLMs providing useful context and common-sense aware reasoning may also be a future possible extension. For instance, prompting LLMs to decompose a long-horizon task into smaller more approachable steps, may not only automate the task decomposition process, yet may also lead to useful emergent behaviour, especially if coupled with IGE-LLMs. Other identified aspects that would be invaluable to pursue as a future advancement to the architecture would be increased higher-dimensionality problems, including multi-expert hierarchical learning as well as tasks requiring bi-manual manipulation. Employing more complex, anthropomorphic robotic hands as end-effectors would also be valuable in future work, likely leading to further closing the gap between the capabilities of humans and AI. Furthermore, leveraging immersive technologies such as MR (as seen in Chapter 3 and Chapter 4), would be promising means of providing demonstration data and rendering the imitation of complex tasks for human demonstrators potentially notably easier than conventional input devices, (Triantafyllidis et al., 2021; Triantafyllidis et al., 2020). Nevertheless, ROMAN's novel architecture is a first step towards underlining the value of deriving biologically-inspired embodied intelligence and fusing these with hierarchical task learning, tailored for notably intricate and long-horizon tasks.

Chapter 6 presented the final contribution of this research thesis on the use of language-guided exploration, leading to the derivation of the IGE-LLMs framework. As with all methods utilising LLMs, the main limitation is the necessity of providing high-level textual scene descriptions. IGE-LLMs being no different, requires input in the form of

textual descriptions. Nevertheless, emerging multimodal LLM solutions or methods such as Contrastive Language-Image Pre-Training (CLIP) (Radford et al., 2021), could automate the input prompts by providing textual descriptions of the exteroceptive states. Combining this with vision could further minimise the Sim2Real gap given its applications to robotic manipulation (Triantafyllidis et al., 2023b). Furthermore, combining the proposed framework with methods such as those proposed by Yu et al. (2023), could aid in bridging high-level instructions to lower-level robotic actions with LLMs. For instance, tasking the LLM to infer and provide low-level auxiliary rewards for task-specific challenges may be another alternative to harnessing their context-aware language insights. Furthermore, IGE-LLMs utilised the powerful *GPT-4* LLM. A comparison of other LLMs or even smaller language models, in general, would aid in a more comprehensive understanding of the framework's robustness. Nevertheless, results showed that even a very capable LLM such as the *GPT-4* model was overall inadequate to be used in isolation, even when employing state-of-the-art reasoning CoT (Wei et al., 2023). This limitation of utilising LLMs as main policy drivers highlights in turn the value of IGE-LLMs and the framework's main advantage of utilising LLMs as an assistive guidance instead. As LLMs are becoming increasingly more capable at inferring abstract reasoning and scene understandings, perhaps including these directly in the decision-making in the near or far future may be a viable option. Lastly, for a more comprehensive sensitivity analysis, different levels of weight decay (w^i) alongside the scaling parameters (λ^i) in this thesis could be evaluated.

Closing Remarks

Overall, while there are numerous frontiers and challenges ahead, the author of this thesis believes that the importance of the fusion of biology and machines may very well be the pivotal multi-disciplinary research necessitated. This, in turn, may lead to experiencing embodied intelligence that may well exceed human capabilities with the central aim of offloading repetitive, strenuous or dangerous tasks from humans altogether. Nevertheless, while an ambitious future lies ahead, this doctoral research presented a blueprint towards the endeavour of closing the gap between humans and machines, to develop machine intelligence inspired from a biological standpoint to solve intricate tasks in robotics. Further advancements in Human Factors, Machine Learning and Robotics are necessitated and the author of this thesis hopes that this work will aid in further bridging and solidifying the gap between humans and AI, to develop closer to biology, embodied intelligence.

Appendix A

Additional Materials – ROMAN

In this appendix, additional technical details of the hierarchical architecture of ROMAN are provided that were not deemed critical to the main understanding of the architecture. Instead, details in this chapter are more elaborate than those presented in the main body of the thesis. The details included in this appendix include the specific hyperparameter settings for training, additional expansion on the results that complement the ones in the main text from Chapter 5 as well as supplementary analyses.

Foremost, additional expansion on the system apparatus as well as simulation overview is given in Section A.1. Next, a thoroughly detailed expansion of the tables and figures from the main thesis is presented in Section A.2, which were however not critical to the main understanding or conclusions of this research. Furthermore, the results of ROMAN are further inferred in Section A.3, complementing existing findings and providing additional details. Lastly, in Section A.4, the employed training procedures for both ROMAN and those utilised in the baseline evaluation of monolithic NN equivalents are listed, with their full NN structures, hyperparameter and settings (states and actions).

A.1 Expansion on the System Overview

To allow the derivation of ROMAN and its subsequent evaluation, the simulation engine Unity3D (v2019.3.0f6) was used. The simulation engine contains the built-in NVIDIA physics engine (PhysX 4.1) coupled with the PyTorch-based ML-Agents toolkit [1], consult Juliani et al. (2018). This, in turn, allowed the simulation to achieve realistic, high-

[1] https://github.com/Unity-Technologies/ml-agents

fidelity physics and ensure that the approximation of a real-life robotics manipulation environment was minimised as much as possible. To further ensure reliable physical modelling of the robotic system (Franka Emika), ROS, and in particular, the ROS# plugin was incorporated to import physics models of robots and objects (URDF) into the Unity3D engine.

In order to further ensure that the physics engine (PhysX 4.1) was as realistic as possible, the physics simulation frequency was set at 1,000Hz. While computationally more demanding, this simulation frequency subsequently ensured robust and stable physics performance entailing realistic physical properties and frictions one would expect from the real world (Triantafyllidis et al., 2021; Yang et al., 2020). It was considered importation to ensure stable and realistic physics performance as combining deep RL methods with realistic simulators is advantageous for training intelligent agents, which is supported by Lee et al. (2019); Yang et al. (2020). This is particularly due to the large number of iterations that can be performed in the simulation contrary to a real-life case, subsequently allowing for a notably easier and quicker interaction of the agent and its environment, while also minimising the risk of physical and hardware damage (Triantafyllidis et al., 2021; Lee et al., 2019; Yang et al., 2020).

The robotic system used for the experiments was the Franka Panda Emika robot, with its default gripper configuration. The evaluation environment, entailing all physical objects manipulated in the simulated tasks was within the robot's workspace. The control policies of the expert NNs, operated in direct velocity control of the end-effector gripper, including the action of opening or closing the gripper. An IK solver was used to solve the kinematic chain to the end-effector position. The IK solver and the control of the robot arm run at 1,000Hz as its default control mode. Moreover, a low-pass filter was introduced and incorporated into the exteroceptive states of all NNs for filtering and de-noising the input signals (i.e. the observation signals) to the NNs. Each expert NN was running at a close-loop frequency of 25Hz to update the end-effector velocity command, while the MN operated at 10Hz loop. Contrary to the expert NNs, the MN outputs consisted of weight assignments to the incorporated expert NNs in the hierarchical architecture. Subsequently, the MN was responsible for the correct and in-sequence coordination of the specialised skills at its disposal, i.e. the expert NNs.

The training of each component of ROMAN was conducted on a desktop computer, with the full training processes on a Central Processing Unit (CPU). The CPU used for this

process was an 18-core Intel i9-9980XE. The parallelism available in the PyTorch-based ML-Agents toolkit, allowed the training procedure to run in concurrent training instances, notably speeding up wall-clock training times.

A.2 Expansion on the Figures and Tables

In this section of the supplementary materials, additional expansion of the tables from the main body of the thesis are presented and in particular the success rates of each sub-task within the full sequence of actions. In regards to the figures, the normalised reward plots from training the expert NNs and the MN, including the baseline evaluation of the monolithic NN are presented. The analyses of the more detailed data allowed for a more comprehensive understanding of not only the overall sequential success rates but also the individual sub-tasks ***within*** the full sequential task decomposition. This subsequently allowed the investigation of whether sequential failures were attributed to either specific expert NN error or due to higher-level supervision errors in the orchestration of weight assignments by the MN.

Expansion on Studied Levels of Uncertainty – In regards to the studied levels of uncertainty, the overall success rates of ROMAN based on different Gaussian levels of exteroceptive noise are detailed in Table A.1, expanding on Table 5.6 in the main body of the thesis. The comparison of the hierarchical formation of ROMAN compared against monolithic NNs equivalents for 2D and 3D, sharing an identical hybrid learning procedure, are detailed in Table A.2 and Table A.3, expanding upon Table 5.4 and Table 5.5 respectively.

Expansion on Ablation Study of Learning Paradigms – In regards to the ablation study and the comparison of the different learning paradigms employed (RL, BC and GAIL), their different combinations and pairing within ROMAN against noise levels equivalent to ± 0.5 cm are detailed in Table A.4 and for ± 1.0 cm as well as ± 2.0 cm in Table A.5, both of which expand upon Table 5.7. For the different number of demonstrations provided to ROMAN and by the extent to its learning paradigms for supervised (BC) and imitation learning (GAIL), the results are presented in Table A.6, expanding upon Table 5.8.

Manipulation Task Completion Times – The duration, in seconds, to complete all case scenarios as well as the sub-tasks within the sequence of actions, by ROMAN against different levels of noise ranging from ±0.0 cm to ±2.5 cm, in 0.5 cm increments, are outlined in Table A.7 for each individual expert and in Table A.8 for the full sequential scenario tasks respectively.

Normalised Reward Plots of NN Training – In Figure A.2 as well as Figure A.3, the maximum attainable reward plots during training are shown for each individual expert within ROMAN and how the different number of demonstrations affected the training procedure ($N = 7$, $N = 21$ and $N = 42$) respectively. Finally, in Figure A.4 the normalised reward plot of a monolithic NN compared against ROMAN is depicted, both sharing an identical hybrid learning procedure, NN settings and hyperparameters in full 3D space. Reward plots are normalised from 0 to 1, over the total time-steps in millions, i.e. $1e7$ or $1e8$, where specified.

A.3 Expansion on the Results

This section provides additional details from the experiments of ROMAN and all employed baseline evaluations and comparisons. These were not deemed critical for the main understanding nor for the conclusions drawn in the main body of the thesis and in particular Chapter 5. Consequently, while these are omitted from the main body, these nonetheless complement existing results and their inclusion as an Appendix was deemed important.

Results Expansion on a Monolithic NN vs ROMAN in 2D. Prior to scaling ROMAN to the more complex 3D Euclidean space composed of seven in total experts, a preliminary benchmark was conducted to evaluate the performance of a monolithic approach versus ROMAN's preliminary version operating in a planar, 2D setting composed of five in total experts. By sharing an identical hybrid learning procedure, this evaluation allowed a direct comparison of a monolithic versus a hierarchical approach. This in turn allowed for the evaluation and demonstration of the advantages of a hierarchical task decomposition. Identical states and hyperparameters were used for both ROMAN's MN and the monolithic NN. The action space of the monolithic NN was identical to ROMAN's experts for controlling the end-effector and the gripper. To further conduct a fair compar-

ison, a total number of $N = 100$ demonstrations were provided to the monolithic single NN, to match ROMAN's 2D setting composed of five experts pre-trained with $N = 20$ each (5 experts $\cdot$ 20 demos each). Table A.2 depicts that while a monolithic NN can accommodate the least complex and with shorter time-horizon sequential scenario cases 1 and 2 well, there is a significant drop in success rates from case 2 to case 3 ($\approx 15\%$). On further investigation, during sub-task completion, it can be observed that the addition of the *Pick and Insert* sub-task exhibits lower success rates than the *Lift* and *Pushing* sub-tasks. In particular, the *Pick and Insert* is the very first sub-task that needs to be activated, consequently, this implies that errors originating from this sub-task will sum up to further sub-task that **directly depend** on the successful completion of the previous sub-task. From Table A.2 it is evident that the **Pick and Insert** sub-task is arguably the most complex one, which is reflected in its individual success rates under that particular column. It is furthermore shown that when expanding to more complex sequences such as cases 4 and 5, the monolithic NN is unable to attain high success rates (lower than $\approx$ 60%). On the other hand, ROMAN exhibits high success rates across cases 1 to 5, attaining at least $\approx 97\%$ success throughout, which underlines the necessity of a hierarchical task decomposition and the overall value of ROMAN's architecture for solving complex and long-horizon in-sequence manipulation tasks.

Results Expansion on a Monolithic NN vs ROMAN in 3D. Similarly to the 2D case of ROMAN, a monolithic NN was again evaluated in the more complex 3D setting composed of a total of seven distinct sub-tasks. Consistent with the baseline evaluation of the monolithic NN versus ROMAN in the 2D case, an identical training procedure via the hybrid learning approach (Figure A.1) was employed for the 3D monolithic NN, with identical states (to ROMAN's MN), actions (to ROMAN's experts) and hyperparameters. For the 3D case, a total of $N = 140$ demonstrations were provided to the monolithic NN to match with ROMAN's seven experts, each pre-trained with $N = 20$ demonstrations (7 experts $\times$ 20 demos). In line with Table A.2, results in Table A.3 suggest that a monolithic NN is unable to solve the complex nature and long sequential task of the validated manipulation scenario given the same training procedure. Although the monolithic NN approach showed robustness against cases S1 to S3 in 2D space, the increased complexity and dimensionality in 3D shows that a monolithic NN significantly struggles with solving more complex sequences. In particular, it is once again shown that a monolithic NN is able to achieve robust results for S1 and S2 (higher than $\approx 98\%$). However, the addition of the

Pick and Insert sub-task in S3, underlines the complexity of the given sub-task, evident in the notably lower success rates under the particular column compared to S1 and S2. Extending to S4 and S5, it is shown that a monolithic NN exhibits almost complete failures (less than $\approx 3\%$ success rates), while entirely unable to complete S6 or S7. Extending the noise level comparison beyond $\sigma = \pm 0.5$ cm was disregarded due to the already notably low success rates of the monolithic NN at such noise levels. It is hence inferred that both in 2D and especially in 3D, a monolithic NN, even when trained with an identical hybrid learning procedure as with ROMAN's incorporated NNs, is unable to solve complex, long-time-horizon sequential manipulation tasks. This implies and highlights the value of the proposed hierarchical approach and the employed task decomposition.

Results Expansion on Increasing Levels of Uncertainty and Ablation of Learning Paradigms. As mentioned in the main body of this thesis and in particular Chapter 5, ROMAN's experts were provided with exteroceptive information relevant to their specific task goals and specialised skills, while the MN used the combined state space of the experts to oversee the context of the environment and the long-term sequential task. ROMAN was later evaluated on a vision-based detection system. However, prior to proceeding to a vision-based system, it was crucial to initially evaluate the architecture beyond the exteroceptive uncertainties the vision system alone would yield. In regards to increasing levels of uncertainty, it was observed that even a five-fold increase in Gaussian noise for exteroceptive states on ROMAN ($\sigma = \pm 2.5$cm), rendered high success rates, as supported by Table A.1. The lowest success rate at that noise level for S7 was 76.2% for the **Pick and Insert** expert which was arguably the most complex of all. It is also worth mentioning that the employed experts were decomposed at a higher-level operating task, in terms of manipulation skills when compared to related works of Marzari et al. (2021); Fox et al. (2019); Rajeswaran et al. (2018). For instance, **Picking and Placing** in the related work of Marzari et al. (2021), was solved using a hierarchical control framework decomposed as (i) *approaching*, (ii) *grasping*, and (iii) *retraction*. Contrarily, the ROMAN architecture supports the capability of deriving the aforementioned as a **single expert**. This in turn allowed the MN to focus on more important higher-level scene understandings of complex task sequences as shown in ROMAN, instead of being focused on orchestrating lower-level skills that can be represented and effectively managed by a single specialising manipulation expert. In regards to the increasing levels of noise for the gating network (the MN), it was observed that by virtue of training ROMAN's MN with

the employed hybrid learning approach detailed in the main body of the thesis, resulted in overall higher and more robust performance than when comparing their different combinations. This implies that an HHL approach exhibits notably higher resilience against (i) increased uncertainty, (ii) the presence of more intricate non-interrelated sub-tasks, (iii) longer time-horizon sequential tasks, as well as (iv) the adaptation to cases beyond those encountered in the demonstration sequence and the ability to dynamically recover from local minima.

Result Expansion on Demonstrations. From the results based on the different number of demonstrations, it was observed that even a relatively small amount of demonstrations for the MN allowed the architecture to attain fairly high success rates even when presented with the most intricate sequences of S6 and S7. Upon detailed evaluation of the results in Table A.6, it was observed that the employed HHL approach (BC + GAIL (r^i) + RL (r^e)) allowed ROMAN to benefit from the initial warm-starting of the policy based on BC, to train a complex gating network with a small number of demonstrations due to the intrinsic reward provided by GAIL, while also retaining robust performance due to the extrinsic RL reward. Overall, from the results, it is initially inferred that a one-shot demonstration of each sequential case scenario ($N = 7$) was not sufficient to solve scenario levels S4 to S7 but still retained success levels of more than $\approx 68\%$ for S1 to S3. Instead, it was observed that for $N = 21$ demonstrations, corresponding to three demonstrations for every scenario, stable and significantly higher success rates were achieved compared to $N = 7$, which were almost at the same level as that of $N = 42$. Even though ROMAN and the employed baseline evaluations including the ablation results were tested at $N = 42$ for most of the comparisons in the main body of the thesis, it is shown that by providing half of those demonstrations ($N = 21$) to the MN, the ROMAN architecture is still sufficient capable of attain high success rates, even amongst the most complex and long-time horizon sequential tasks as evidenced by Table A.6.

A.4 Expansion on Training and Hyperparameters

In addition to the training details outlined in the main body of the thesis and in particular Chapter 5, in this section, the training procedure employed to train all incorporated NNs within the hierarchical formation of ROMAN is elaborated upon. Figure A.1 visually depicts the hybrid learning procedure in the form of a flow diagram.

ROMAN's Neural Network Settings and Hyperparameters. Table A.9 reports ROMAN's preliminary stage composed of five in total experts in 2D space, detailing the states and action space of all incorporated NNs in the hierarchy. Table A.10 depicts the architecture overview of ROMAN's final stage in 3D space composed of seven experts, including the details of the states, actions, demonstrations and dimensions overview of all incorporated NNs, experts and MN, in the hierarchical architecture. Furthermore, the hyperparameters used for the hybrid hierarchical learning procedure of ROMAN's NNs are listed in Table A.11.

Monolithic Neural Network Settings and Hyperparameters. For the two monolithic NNs in 2D and 3D space used as part of the baseline evaluation against ROMAN's hierarchical architecture, their architectural settings, details and hyperparameters are listed. Table A.12 depicts the architectures of the monolithic NNs in 2D and 3D space, reporting their states, actions, demonstrations and NN architectures. An identical hybrid learning procedure was used to train the monolithic NNs and ROMAN to adhere to consistency for subsequent fair comparisons. Moreover, identical hyperparameter values for the monolithic NNs as for ROMAN's MN, in 2D and 3D respectively were utilised. The specific hyperparameters used for the monolithic NNs in 2D and 3D are listed in Table A.13. Consequently, this translated and allowed for a fair comparison of a monolithic vs hierarchical approach and ultimately underlined the necessity and usefulness of a hierarchical task decomposition as with ROMAN's architecture.

Reward Design. In regards to the reward design of the expert NNs, the employed rewards were specific to their sub-tasks and the nature of their objectives, through a combination of sparse rewards with a terminal reward at the end of a successful episode. The gating network (the MN) was awarded by successfully completing each sub-task, in addition to a terminal reward for successfully completing the entire sequence of tasks.

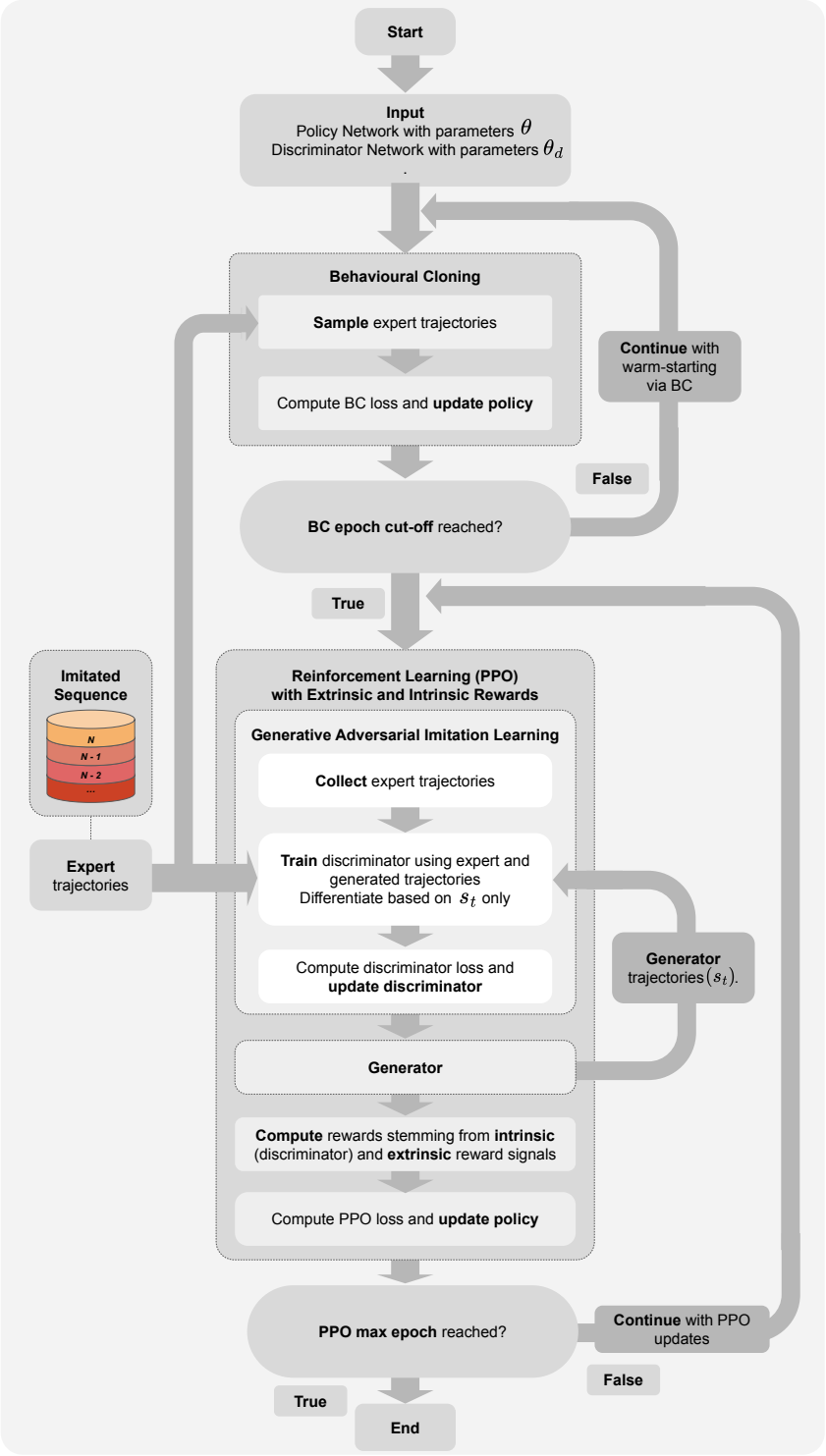

Figure A.1: **Flow chart of the hybrid training procedure.** The figure depicts the main stages of the training procedure, including the use of demonstrations to warm-start the policy via BC, followed by the use of PPO to update the policy with extrinsic (r^e) and intrinsic (r^i) rewards provided by the environment and GAIL's discriminator respectively. This training procedure is repeated for all expert NNs incorporated in ROMAN's hierarchical architecture. Given the pre-trained expert NNs, the MN is then trained with the same hybrid learning procedure.

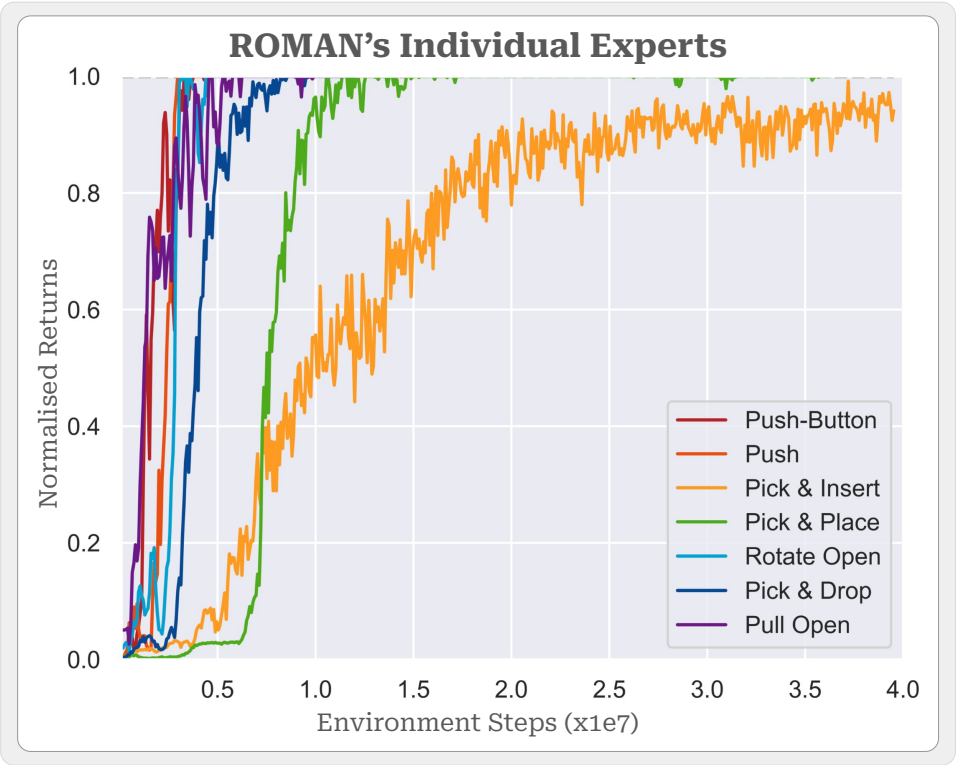

Figure A.2: **Returns plot for each expert NN within ROMAN's architecture, depicting the normalised reward over the environment steps in millions.** The figure shows the environment turns over the environment steps for each of the seven in total expert NNs in ROMAN's hierarchical learning architecture. Notice that the training requirements in environment steps depend on the nature and complexity of each specialising expert, corresponding, in turn, to quicker or slower convergence towards the maximum attainable reward. For instance, expert NNs concerned with higher-level complexity goals, such as those concerned with *Picking & Dropping, Placing or Inserting*, were the most intricate and longest in time horizons compared to others. As discussed in Chapter 5, developing task-specific specialising experts allowed the off-loading of secondary information from the gating network (the MN) and instead allowed it to focus entirely on inferring higher-level scene understandings and the plethora of specialised experts in the architecture. This approach minimised the amount of secondary information that the gating network needs to process, resulting in more efficient and effective task execution. As inferred in the returns plot, the highest in complexity expert was the *Picking and Inserting*, necessitating the most training in environment steps compared to other experts. This was in line with the qualitative difficulty of obtaining the demonstration data from the human expert, which was also more demanding in this specific sub-task. All expert NNs were pre-trained with a total of $N = 20$ demonstrations.

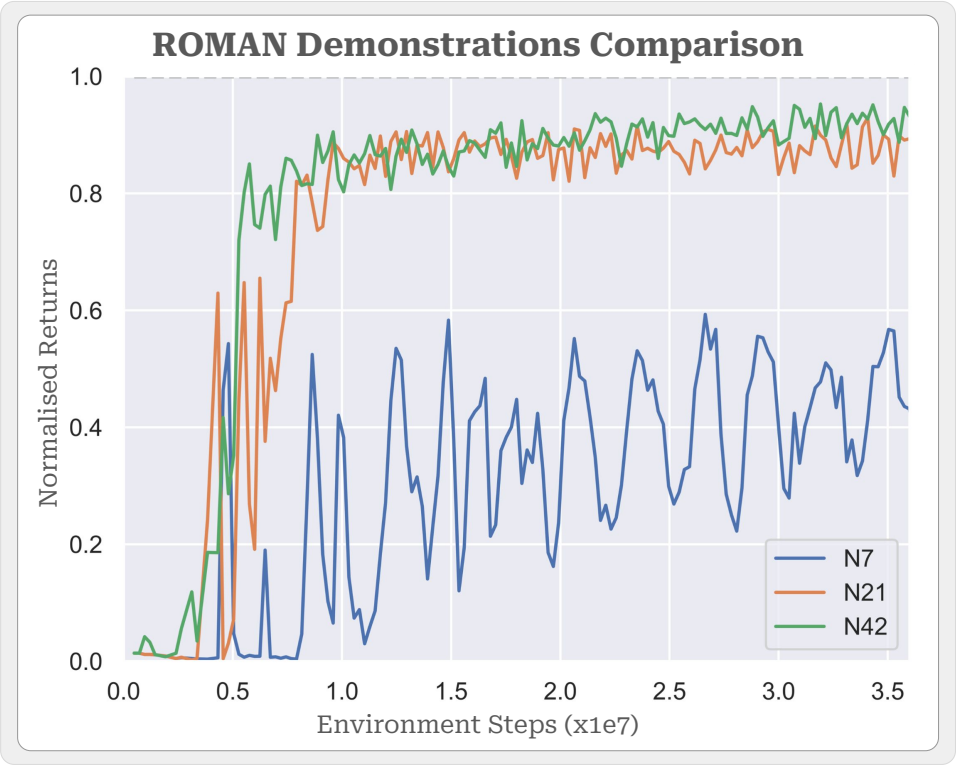

Figure A.3: **Returns plot for the different number of demonstrations on the ROMAN architecture, with the normalised reward over the environment steps in millions.** The figure depicts the effects of the different number of demonstrations ($N = 7$, $N = 21$ and $N = 42$) on ROMAN's gating network (the MN) and how it affected training. As described in more detail in Chapter 5, a total of $N = 7$ demonstrations, corresponded essentially to one demonstration for each of the seven different sequential scenario tasks. Similarly, $N = 21$ and $N = 42$ correspond to three and six demos per scenario respectively. After analysing the above reward plot and testing ROMAN on these different numbers of demonstrations, it is concluded that $N = 7$ was insufficient for a minimal level of acceptable performance of the gating network. Instead, a minimum of $N = 21$ demonstrations were necessary to attain close to the maximum returns.

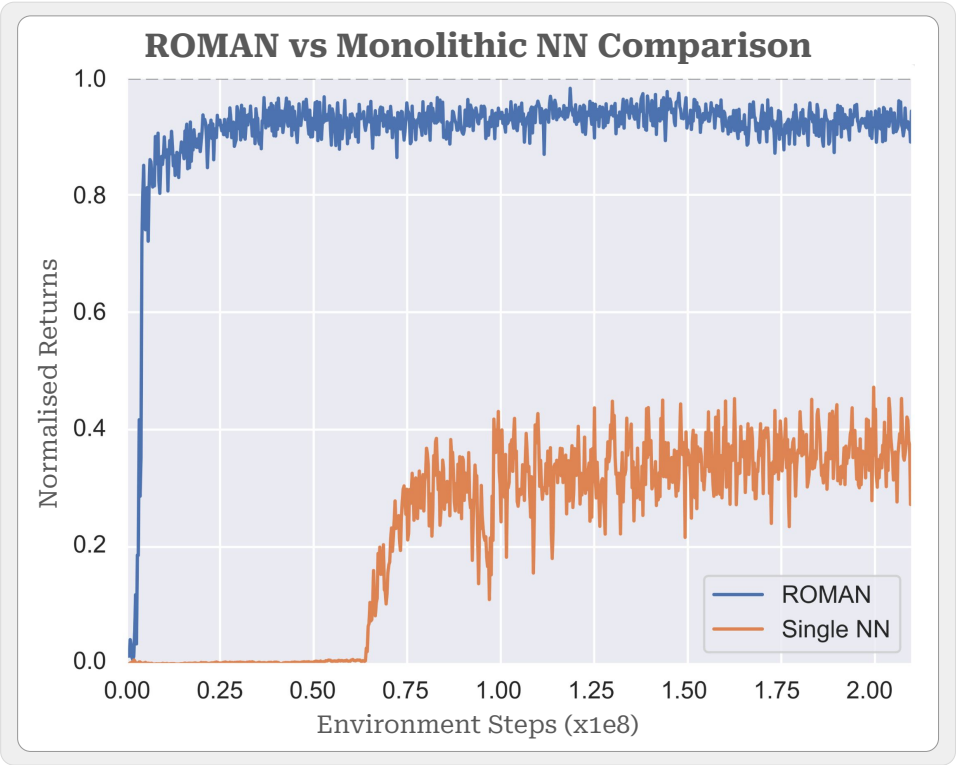

Figure A.4: **Returns plot between comparing the value of a hierarchical architecture: ROMAN versus a monolithic NN in full 3D space, with the overall normalised returns over the environment steps in tens of millions.** The figure depicts the normalised returns plot of ROMAN vs a monolithic NN, both being trained in full 3D space. It is inferred that only ROMAN is able to converge close to the maximum attainable reward. ROMAN's MN was provided with a total of $N = 42$ demonstrations, corresponding to $N = 6$ demonstrations for each of the seven sequential case scenarios. To conduct a fair comparison, the monolithic NN in 3D was given $N = 140$ demonstrations in total to account for ROMAN's seven pre-trained experts with $N = 20$ demonstrations each. To further retain consistency, it is noteworthy to point out that both ROMAN (and all entailed NNs therein) as well as the monolithic NN, were trained with an identical hybrid learning procedure. The returns plot suggests that a hierarchical formation is of necessity for intricate sequential and long-time horizon tasks as studied in Chapter 5. Based on the returns plot and the results inferred in Section 5.3, a monolithic NN, even when provided access to the same number of demonstrations and trained with an identical hybrid learning procedure, is unable to attain robust performance. In conclusion, this highlights the value of hierarchical task decomposition, in order to address notably intricate long-horizon sequential tasks as commonly seen in robotic manipulation.

Table A.1: **Results of ROMAN expanding further upon Table 5.6, where the overall success rates of each sequential case scenario are depicted, across different levels of uncertainty, with the addition of the individual expert success highlighted.** The summary of the overall success rates is shown with each sequential case scenario over different levels of uncertainty. Contrary to Table 5.6, this table also illustrates the addition of the individual expert success rate sequences within the action sequence for the particular sequential case. The grey cells, denoted as n/a, are omitted in those sequential case scenarios where those experts are not part of the sequential actions needed to satisfy the end-goal task. In particular, only cells that are relevant to the scenario are studied, the rest are omitted. For an overall sequential case scenario to be deemed successful, all relevant sub-tasks needed to be satisfied and in a particular order. The MN was provided with a total of $N = 42$ demonstrations, which corresponds to $N = 6$ per each of the seven case scenarios. Each cell stems from 1,000 trials.

Detailed Success Rates of ROMAN Against Different Levels of Uncertainty									
	Scenario Cases	**Push-Button**	**Push**	**Pick & Insert**	**Pick & Place**	**Rotate Open**	**Pick & Drop**	**Pull-Open**	**Overall**
$\sigma = \pm 0.0$ [cm]	**S1:** *Push-Button*	0.976	n/a	n/a	n/a	n/a	n/a	n/a	0.976
	S2: *+Push*	0.972	0.991	n/a	n/a	n/a	n/a	n/a	0.972
	S3: *+Pick & Insert*	0.869	0.876	0.881	n/a	n/a	n/a	n/a	0.847
	S4: *+Pick & Place*	0.955	0.975	0.971	0.982	n/a	n/a	n/a	0.951
	S5: *+Rotate Open*	0.743	0.750	0.753	0.768	0.771	n/a	n/a	0.728
	S6: *+Pick & Drop*	0.961	0.968	0.965	0.976	0.999	0.993	n/a	0.954
	S7: *+Pull-Open*	0.912	0.924	0.917	0.942	0.972	0.973	0.972	0.903
$\sigma = \pm 0.5$ [cm]	**S1:** *Push-Button*	0.973	n/a	n/a	n/a	n/a	n/a	n/a	0.973
	S2: *+Push*	0.975	0.989	n/a	n/a	n/a	n/a	n/a	0.975
	S3: *+Pick & Insert*	0.841	0.854	0.869	n/a	n/a	n/a	n/a	0.817
	S4: *+Pick & Place*	0.965	0.966	0.966	0.985	n/a	n/a	n/a	0.959
	S5: *+Rotate Open*	0.811	0.822	0.820	0.835	0.833	n/a	n/a	0.794
	S6: *+Pick & Drop*	0.967	0.973	0.971	0.983	0.987	0.992	n/a	0.960
	S7: *+Pull-Open*	0.958	0.964	0.962	0.983	0.991	0.993	0.994	0.952
$\sigma = \pm 1.0$ [cm]	**S1:** *Push-Button*	0.977	n/a	n/a	n/a	n/a	n/a	n/a	0.977
	S2: *+Push*	0.999	0.997	n/a	n/a	n/a	n/a	n/a	0.990
	S3: *+Pick & Insert*	0.831	0.841	0.843	n/a	n/a	n/a	n/a	0.798
	S4: *+Pick & Place*	0.955	0.960	0.955	0.980	n/a	n/a	n/a	0.946
	S5: *+Rotate Open*	0.796	0.802	0.799	0.813	0.816	n/a	n/a	0.776
	S6: *+Pick & Drop*	0.952	0.955	0.946	0.956	0.974	0.984	n/a	0.933
	S7: *+Pull-Open*	0.954	0.955	0.944	0.970	0.987	0.999	0.989	0.939
$\sigma = \pm 1.5$ [cm]	**S1:** *Push-Button*	0.980	n/a	n/a	n/a	n/a	n/a	n/a	0.980
	S2: *+Push*	0.986	0.996	n/a	n/a	n/a	n/a	n/a	0.986
	S3: *+Pick & Insert*	0.756	0.767	0.794	n/a	n/a	n/a	n/a	0.720
	S4: *+Pick & Place*	0.881	0.882	0.874	0.930	n/a	n/a	n/a	0.846
	S5: *+Rotate Open*	0.767	0.766	0.758	0.808	0.853	n/a	n/a	0.722
	S6: *+Pick & Drop*	0.875	0.870	0.860	0.912	0.973	0.978	n/a	0.836
	S7: *+Pull-Open*	0.875	0.871	0.862	0.933	0.979	0.983	0.986	0.841
$\sigma = \pm 2.0$ [cm]	**S1:** *Push-Button*	0.967	n/a	n/a	n/a	n/a	n/a	n/a	0.967
	S2: *+Push*	0.986	0.994	n/a	n/a	n/a	n/a	n/a	0.986
	S3: *+Pick & Insert*	0.768	0.783	0.805	n/a	n/a	n/a	n/a	0.737
	S4: *+Pick & Place*	0.862	0.865	0.859	0.901	n/a	n/a	n/a	0.837
	S5: *+Rotate Open*	0.785	0.790	0.790	0.829	0.892	n/a	n/a	0.753
	S6: *+Pick & Drop*	0.842	0.843	0.850	0.882	0.943	0.954	n/a	0.820
	S7: *+Pull-Open*	0.840	0.839	0.840	0.886	0.973	0.977	0.983	0.815
$\sigma = \pm 2.5$ [cm]	**S1:** *Push-Button*	0.973	n/a	n/a	n/a	n/a	n/a	n/a	0.973
	S2: *+Push*	0.986	0.995	n/a	n/a	n/a	n/a	n/a	0.986
	S3: *+Pick & Insert*	0.767	0.774	0.778	n/a	n/a	n/a	n/a	0.723
	S4: *+Pick & Place*	0.789	0.790	0.800	0.856	n/a	n/a	n/a	0.763
	S5: *+Rotate Open*	0.740	0.748	0.739	0.784	0.898	n/a	n/a	0.697
	S6: *+Pick & Drop*	0.751	0.749	0.769	0.807	0.931	0.935	n/a	0.719
	S7: *+Pull-Open*	0.788	0.785	0.762	0.830	0.962	0.964	0.980	0.744

Table A.2: **Detailed results of a monolithic NN vs the preliminary stage of the hierarchical architecture ROMAN in 2D composed of five experts, expanding upon Table 5.4.** This table summarises the percentage of successful sequential scenarios over the five in total decomposed sub-tasks for ROMAN's preliminary version. Consequently, the sub-tasks listed under the five different columns are perceived here as different sub-tasks within the time horizon of the full sequence of the episode. This addition of the sub-task success rates within the sequence of actions allowed for a more in-depth analysis. A total of $N = 35$ demonstrations were provided to the MN of ROMAN, while for the monolithic NN, a total of $N = 100$ demonstrations were given to conduct a fair comparison and to account for ROMAN's five experts in total that were pre-trained with $N = 20$ demonstrations ($5_{\text{experts}} \cdot 20_{\text{demos}}$). Identical hyperparameters were used, with the states of the monolithic NN corresponding to that of the MN's, while the actions of the monolithic NN were identical expert NNs of ROMAN's. Evaluated on $\sigma = \pm 0.5$ [cm] level of noise. Each cell stems from 1,000 trials.

[2D] Monolithic Single NN vs Preliminary Stage of the Hierarchical architecture ROMAN							
	Scenario Cases	**Push**	**Lift**	**Pick & Insert**	**Pick & Drop**	**Pull**	**Overall**
Single NN	**S1:** *Push*	0.997	n/a	n/a	n/a	n/a	0.997
	S2: *+Lift*	0.851	0.987	n/a	n/a	n/a	0.841
	S3: *+Pick & Insert*	0.843	0.945	0.823	n/a	n/a	0.699
	S4: *+Pick & Drop*	0.702	0.782	0.679	0.869	n/a	0.591
	S5: *+Pull*	0.718	0.832	0.700	0.919	0.987	0.565
ROMAN	**S1:** *Push*	0.993	n/a	n/a	n/a	n/a	0.993
	S2: *+Lift*	0.996	0.997	n/a	n/a	n/a	0.995
	S3: *+Pick & Insert*	0.996	0.997	0.987	n/a	n/a	0.982
	S4: *+Pick & Drop*	0.977	0.981	0.978	0.991	n/a	0.971
	S5: *+Pull*	0.985	0.986	0.986	0.993	0.998	0.974

Table A.3: **Detailed results of a monolithic NN vs the hierarchical architecture of ROMAN in its final stage in 3D composed of seven experts, expanding upon Table 5.5.** Similarly to Table A.2, this table represents the percentage of successful sequential scenarios for each decomposed subtask for ROMAN's final stage. ROMAN's MN was provided $N = 42$ demonstrations, while the monolithic NN in 3D was provided a total of $N = 140$, to account for ROMAN's seven in total experts pre-trained with $N = 20$ demonstrations each ($7_{\text{experts}} \cdot 20_{\text{demos}}$). An identical hybrid learning procedure as outlined in the main thesis was employed, with identical states (to ROMAN's MN), actions (to ROMAN's expert NNs) and hyperparameters. Evaluated on $\sigma = \pm 0.5$ [cm] level of noise. Each cell stems from 1,000 trials.

[3D] Monolithic Single NN vs Hierarchical Architecture ROMAN									
	Scenario Cases	**Push-Button**	**Push**	**Pick & Insert**	**Pick & Place**	**Rotate Open**	**Pick & Drop**	**Pull-Open**	**Overall**
Single NN (±0.5 cm)	**S1:** *Push-Button*	0.997	n/a	n/a	n/a	n/a	n/a	n/a	0.997
	S2: *+Push*	0.982	0.987	n/a	n/a	n/a	n/a	n/a	0.981
	S3: *+Pick & Insert*	0.671	0.678	0.587	n/a	n/a	n/a	n/a	0.583
	S4: *+Pick & Place*	0.041	0.046	0.036	0.103	n/a	n/a	n/a	0.032
	S5: *+Rotate Open*	0.041	0.044	0.035	0.076	0.934	n/a	n/a	0.028
	S6: *+Pick & Drop*	0.000	0.000	0.000	0.006	0.006	0.024	n/a	0.000
	S7: *+Pull-Open*	0.000	0.000	0.000	0.000	0.010	0.028	0.877	0.000
ROMAN (±0.5 cm)	**S1:** *Push-Button*	0.973	n/a	n/a	n/a	n/a	n/a	n/a	0.973
	S2: *+Push*	0.975	0.989	n/a	n/a	n/a	n/a	n/a	0.975
	S3: *+Pick & Insert*	0.841	0.854	0.869	n/a	n/a	n/a	n/a	0.817
	S4: *+Pick & Place*	0.965	0.966	0.966	0.985	n/a	n/a	n/a	0.959
	S5: *+Rotate Open*	0.882	0.892	0.892	0.938	0.949	n/a	n/a	0.852
	S6: *+Pick & Drop*	0.967	0.973	0.971	0.983	0.987	0.992	n/a	0.960
	S7: *+Pull-Open*	0.958	0.964	0.962	0.983	0.991	0.993	0.994	0.952

Table A.4: **Ablation results of ROMAN's employed learning paradigms, expanding upon Table 5.7 for $\sigma = \pm 0.5$ [cm] level of noise.** The table summarises the overall success rates of each sequential scenario, in addition to the individual expert success rates based on different combinations and pairs of BC, RL and GAIL. **Notation on BC:** Supervised learning on the dataset. **Notation on GAIL:** Use of IL intrinsic rewards (r^i) provided to PPO. **Notation on RL:** Use of task extrinsic rewards (r^e) provided to PPO. **Notation of ROMAN's †:** Default HHL approach combining BC, IL (via r^i) and RL (via r^e). Each cell stems from 1,000 trials.

Detailed Success Rates of Different Algorithms within ROMAN at Noise Level: ±0.5 cm									
	Scenario Cases	**Push-Button**	**Push**	**Pick & Insert**	**Pick & Place**	**Rotate Open**	**Pick & Drop**	**Pull-Open**	**Overall**
RL	**S1:** *Push-Button*	0.000	n/a	n/a	n/a	n/a	n/a	n/a	0.000
	S2: *+Push*	0.000	0.035	n/a	n/a	n/a	n/a	n/a	0.000
	S3: *+Pick & Insert*	0.000	0.000	0.000	n/a	n/a	n/a	n/a	0.000
	S4: *+Pick & Place*	0.000	0.000	0.000	0.000	n/a	n/a	n/a	0.000
	S5: *+Rotate Open*	0.000	0.000	0.000	0.000	0.000	n/a	n/a	0.000
	S6: *+Pick & Drop*	0.000	0.000	0.000	0.000	0.000	0.000	n/a	0.000
	S7: *+Pull-Open*	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GAIL	**S1:** *Push-Button*	0.980	n/a	n/a	n/a	n/a	n/a	n/a	0.980
	S2: *+Push*	0.946	0.493	n/a	n/a	n/a	n/a	n/a	0.468
	S3: *+Pick & Insert*	0.610	0.649	0.708	n/a	n/a	n/a	n/a	0.559
	S4: *+Pick & Place*	0.014	0.013	0.015	0.028	n/a	n/a	n/a	0.012
	S5: *+Rotate Open*	0.005	0.006	0.008	0.034	0.757	n/a	n/a	0.003
	S6: *+Pick & Drop*	0.004	0.005	0.006	0.025	0.755	0.980	n/a	0.001
	S7: *+Pull-Open*	0.003	0.004	0.006	0.020	0.798	0.972	0.847	0.000
BC	**S1:** *Push-Button*	0.986	n/a	n/a	n/a	n/a	n/a	n/a	0.986
	S2: *+Push*	0.979	0.994	n/a	n/a	n/a	n/a	n/a	0.978
	S3: *+Pick & Insert*	0.828	0.844	0.826	n/a	n/a	n/a	n/a	0.786
	S4: *+Pick & Place*	0.694	0.689	0.677	0.830	n/a	n/a	n/a	0.660
	S5: *+Rotate Open*	0.565	0.557	0.546	0.650	0.698	n/a	n/a	0.525
	S6: *+Pick & Drop*	0.768	0.764	0.745	0.887	0.937	0.983	n/a	0.722
	S7: *+Pull-Open*	0.813	0.798	0.776	0.889	0.934	0.958	0.964	0.760
RL, GAIL	**S1:** *Push-Button*	0.981	n/a	n/a	n/a	n/a	n/a	n/a	0.981
	S2: *+Push*	0.942	0.492	n/a	n/a	n/a	n/a	n/a	0.468
	S3: *+Pick & Insert*	0.639	0.663	0.717	n/a	n/a	n/a	n/a	0.570
	S4: *+Pick & Place*	0.010	0.010	0.012	0.025	n/a	n/a	n/a	0.009
	S5: *+Rotate Open*	0.007	0.009	0.009	0.022	0.757	n/a	n/a	0.005
	S6: *+Pick & Drop*	0.015	0.016	0.013	0.024	0.737	0.982	n/a	0.006
	S7: *+Pull-Open*	0.004	0.006	0.005	0.021	0.786	0.968	0.864	0.004
RL, BC	**S1:** *Push-Button*	0.995	n/a	n/a	n/a	n/a	n/a	n/a	0.995
	S2: *+Push*	0.965	0.912	n/a	n/a	n/a	n/a	n/a	0.897
	S3: *+Pick & Insert*	0.907	0.902	0.875	n/a	n/a	n/a	n/a	0.841
	S4: *+Pick & Place*	0.894	0.703	0.689	0.772	n/a	n/a	n/a	0.683
	S5: *+Rotate Open*	0.532	0.509	0.500	0.549	0.572	n/a	n/a	0.492
	S6: *+Pick & Drop*	0.866	0.846	0.831	0.893	0.869	0.956	n/a	0.754
	S7: *+Pull-Open*	0.911	0.866	0.848	0.923	0.892	0.967	0.967	0.774
ROMAN's †	**S1:** *Push-Button*	0.973	n/a	n/a	n/a	n/a	n/a	n/a	0.973
	S2: *+Push*	0.975	0.989	n/a	n/a	n/a	n/a	n/a	0.975
	S3: *+Pick & Insert*	0.841	0.854	0.869	n/a	n/a	n/a	n/a	0.817
	S4: *+Pick & Place*	0.965	0.966	0.966	0.985	n/a	n/a	n/a	0.959
	S5: *+Rotate Open*	0.882	0.892	0.892	0.938	0.949	n/a	n/a	0.852
	S6: *+Pick & Drop*	0.967	0.973	0.971	0.983	0.987	0.992	n/a	0.960
	S7: *+Pull-Open*	0.958	0.964	0.962	0.983	0.991	0.993	0.994	0.952

Table A.5: **Ablation results of ROMAN's employed learning paradigms, expanding upon Table 5.7 for $\sigma = \pm 1.0$ [cm] and $\sigma = \pm 2.0$ [cm] levels of noise.** The summary of the overall success rates is based on different combinations and pairs of BC, RL and GAIL. **Notation on BC:** Supervised learning on the dataset. **Notation on GAIL:** Use of IL intrinsic rewards (r^i) provided to PPO. **Notation on RL:** Use of task extrinsic rewards (r^e) provided to PPO. **Notation on ROMAN's †:** Default HHL approach combining BC, IL (via r^i) and RL (via r^e). Each cell stems from 1,000 trials.

Detailed Success Rates of Different Algorithms within ROMAN at Noise Level: ±1.0 cm									
	Scenario Cases	**Push-Button**	**Push**	**Pick & Insert**	**Pick & Place**	**Rotate Open**	**Pick & Drop**	**Pull-Open**	**Overall**
BC	**S1:** *Push-Button*	0.995	n/a	n/a	n/a	n/a	n/a	n/a	0.995
	S2: *+Push*	0.991	0.996	n/a	n/a	n/a	n/a	n/a	0.990
	S3: *+Pick & Insert*	0.798	0.786	0.740	n/a	n/a	n/a	n/a	0.712
	S4: *+Pick & Place*	0.650	0.629	0.578	0.810	n/a	n/a	n/a	0.573
	S5: *+Rotate Open*	0.543	0.522	0.483	0.655	0.750	n/a	n/a	0.474
	S6: *+Pick & Drop*	0.657	0.606	0.571	0.723	0.804	0.973	n/a	0.563
	S7: *+Pull-Open*	0.748	0.684	0.637	0.823	0.921	0.978	0.992	0.632
RL, BC	**S1:** *Push-Button*	0.996	n/a	n/a	n/a	n/a	n/a	n/a	0.996
	S2: *+Push*	0.990	0.897	n/a	n/a	n/a	n/a	n/a	0.895
	S3: *+Pick & Insert*	0.929	0.923	0.916	n/a	n/a	n/a	n/a	0.881
	S4: *+Pick & Place*	0.941	0.780	0.768	0.811	n/a	n/a	n/a	0.766
	S5: *+Rotate Open*	0.605	0.595	0.586	0.608	0.613	n/a	n/a	0.562
	S6: *+Pick & Drop*	0.839	0.830	0.816	0.849	0.774	0.944	n/a	0.696
	S7: *+Pull-Open*	0.890	0.868	0.863	0.895	0.785	0.976	0.980	0.729
ROMAN's	**S1:** *Push-Button*	0.977	n/a	n/a	n/a	n/a	n/a	n/a	0.977
	S2: *+Push*	0.999	0.997	n/a	n/a	n/a	n/a	n/a	0.990
	S3: *+Pick & Insert*	0.831	0.841	0.843	n/a	n/a	n/a	n/a	0.798
	S4: *+Pick & Place*	0.955	0.960	0.955	0.980	n/a	n/a	n/a	0.946
	S5: *+Rotate Open*	0.796	0.802	0.799	0.813	0.816	n/a	n/a	0.776
	S6: *+Pick & Drop*	0.952	0.955	0.946	0.956	0.974	0.984	n/a	0.933
	S7: *+Pull-Open*	0.954	0.955	0.944	0.970	0.987	0.999	0.989	0.939

(a) Algorithm Comparison at ±1.0 cm of Gaussian Noise.

Detailed Success Rates of Different Algorithms within ROMAN at Noise Level: ±2.0 cm									
	Scenario Cases	**Push-Button**	**Push**	**Pick & Insert**	**Pick & Place**	**Rotate Open**	**Pick & Drop**	**Pull-Open**	**Overall**
BC	**S1:** *Push-Button*	0.838	n/a	n/a	n/a	n/a	n/a	n/a	0.838
	S2: *+Push*	0.678	0.902	n/a	n/a	n/a	n/a	n/a	0.678
	S3: *+Pick & Insert*	0.657	0.732	0.733	n/a	n/a	n/a	n/a	0.609
	S4: *+Pick & Place*	0.222	0.235	0.227	0.338	n/a	n/a	n/a	0.205
	S5: *+Rotate Open*	0.210	0.225	0.216	0.290	0.881	n/a	n/a	0.190
	S6: *+Pick & Drop*	0.124	0.128	0.121	0.163	0.417	0.454	n/a	0.111
	S7: *+Pull-Open*	0.085	0.085	0.081	0.096	0.215	0.229	0.949	0.075
RL, BC	**S1:** *Push-Button*	0.947	n/a	n/a	n/a	n/a	n/a	n/a	0.947
	S2: *+Push*	0.886	0.899	n/a	n/a	n/a	n/a	n/a	0.841
	S3: *+Pick & Insert*	0.817	0.815	0.799	n/a	n/a	n/a	n/a	0.725
	S4: *+Pick & Place*	0.599	0.478	0.456	0.525	n/a	n/a	n/a	0.442
	S5: *+Rotate Open*	0.420	0.393	0.385	0.453	0.718	n/a	n/a	0.363
	S6: *+Pick & Drop*	0.298	0.285	0.282	0.332	0.472	0.518	n/a	0.246
	S7: *+Pull-Open*	0.131	0.121	0.119	0.147	0.214	0.236	0.943	0.100
ROMAN's	**S1:** *Push-Button*	0.967	n/a	n/a	n/a	n/a	n/a	n/a	0.967
	S2: *+Push*	0.986	0.994	n/a	n/a	n/a	n/a	n/a	0.986
	S3: *+Pick & Insert*	0.768	0.783	0.805	n/a	n/a	n/a	n/a	0.737
	S4: *+Pick & Place*	0.862	0.865	0.859	0.901	n/a	n/a	n/a	0.837
	S5: *+Rotate Open*	0.785	0.790	0.790	0.829	0.892	n/a	n/a	0.753
	S6: *+Pick & Drop*	0.842	0.843	0.850	0.882	0.943	0.954	n/a	0.820
	S7: *+Pull-Open*	0.840	0.839	0.840	0.886	0.973	0.977	0.983	0.815

(b) Algorithm Comparison at ±2.0 cm of Gaussian noise.

Table A.6: **Detailed results of ROMAN expanding upon Table 5.8, whereby the overall success of each sequential case scenario is depicted over the different number of demonstrations, with the individual expert success further highlighted.** The table summarises the overall success rates of each sequential case scenario, in addition to the individual expert success rates, based on the different number of demonstrations provided to the gating network (MN) for a total of $N = 7$, $N = 21$ and $N = 42$. It is worth noting that $N = 7$ correspond to one demonstration for each of the seven case scenarios. Hence, $N = 7$, $N = 21$ and $N = 42$ correspond to a total of $N = 1$, $N = 3$ and $N = 6$ demonstrations for each sequential case respectively. Evaluated on $\sigma = \pm 0.5$ [cm] level of noise. Each cell stems from 1,000 trials.

Detailed Success Rates of ROMAN with Different Number of Demonstrations									
	Scenario Cases	**Push-Button**	**Push**	**Pick & Insert**	**Pick & Place**	**Rotate Open**	**Pick & Drop**	**Pull-Open**	**Overall**
$N = 7$	**S1:** *Push-Button*	0.775	n/a	n/a	n/a	n/a	n/a	n/a	0.775
	S2: *+Push*	0.888	0.994	n/a	n/a	n/a	n/a	n/a	0.876
	S3: *+Pick & Insert*	0.732	0.775	0.818	n/a	n/a	n/a	n/a	0.680
	S4: *+Pick & Place*	0.391	0.394	0.409	0.498	n/a	n/a	n/a	0.378
	S5: *+Rotate Open*	0.377	0.386	0.414	0.515	0.970	n/a	n/a	0.360
	S6: *+Pick & Drop*	0.008	0.008	0.008	0.013	0.031	0.033	n/a	0.008
	S7: *+Pull-Open*	0.005	0.009	0.010	0.013	0.033	0.035	0.978	0.005
$N = 21$	**S1:** *Push-Button*	0.994	n/a	n/a	n/a	n/a	n/a	n/a	0.994
	S2: *+Push*	0.923	0.934	n/a	n/a	n/a	n/a	n/a	0.921
	S3: *+Pick & Insert*	0.731	0.745	0.739	n/a	n/a	n/a	n/a	0.718
	S4: *+Pick & Place*	0.948	0.958	0.957	0.986	n/a	n/a	n/a	0.945
	S5: *+Rotate Open*	0.944	0.951	0.945	0.969	0.947	n/a	n/a	0.902
	S6: *+Pick & Drop*	0.935	0.940	0.938	0.957	0.969	0.974	n/a	0.929
	S7: *+Pull-Open*	0.962	0.969	0.969	0.988	0.992	0.995	0.996	0.958
$N = 42$	**S1:** *Push-Button*	0.973	n/a	n/a	n/a	n/a	n/a	n/a	0.973
	S2: *+Push*	0.975	0.989	n/a	n/a	n/a	n/a	n/a	0.975
	S3: *+Pick & Insert*	0.841	0.854	0.869	n/a	n/a	n/a	n/a	0.817
	S4: *+Pick & Place*	0.965	0.966	0.966	0.985	n/a	n/a	n/a	0.959
	S5: *+Rotate Open*	0.882	0.892	0.892	0.938	0.949	n/a	n/a	0.852
	S6: *+Pick & Drop*	0.967	0.973	0.971	0.983	0.987	0.992	n/a	0.960
	S7: *+Pull-Open*	0.958	0.964	0.962	0.983	0.991	0.993	0.994	0.952

Table A.7: **Duration of each expert within ROMAN completing their individual sub-task.** The table reports the average duration, in seconds as well as the standard deviation of each expert completing their specialised sub-task. Results were obtained amongst 10,000 trials per cell at noise levels ranging from $\sigma = \pm 0.0$ [cm] to $\sigma = \pm 2.5$ [cm] with 0.5 [cm] increments. Symbols of $\downarrow$ and $\uparrow$ denote a decrease and increase in time difference compared to the previous noise levels respectively.

The Duration of Individual Expert Sub-Task of ROMAN							
Noise Levels	*Push-Button*	*Push*	*Pick & Insert*	*Pick & Place*	*Rotate Open*	*Pick & Drop*	*Pull-Open*
$\sigma = \pm$**0.0** [cm]	$24.32 \pm 7.55s$	$16.25 \pm 9.03s$	$32.87 \pm 13.89s$	$28.53 \pm 6.77s$	$21.83 \pm 5.80s$	$22.61 \pm 5.05s$	$27.66 \pm 14.34s$
$\sigma = \pm$**0.5** [cm]	$\downarrow 23.69 \pm 6.89s$	$\downarrow 15.72 \pm 8.38s$	$\downarrow 32.15 \pm 12.93s$	$\uparrow 28.61 \pm 6.32s$	$\downarrow 21.48 \pm 4.75s$	$\uparrow 22.67 \pm 5.85s$	$\downarrow 21.18 \pm 5.90s$
$\sigma = \pm$**1.0** [cm]	$\downarrow 22.62 \pm 4.21s$	$\downarrow 15.64 \pm 7.66s$	$\uparrow 32.48 \pm 12.27s$	$\uparrow 29.90 \pm 7.76s$	$\uparrow 21.60 \pm 5.29s$	$\uparrow 23.60 \pm 6.91s$	$\uparrow 21.19 \pm 6.37s$
$\sigma = \pm$**1.5** [cm]	$\downarrow 22.36 \pm 4.05s$	$\uparrow 15.90 \pm 7.41s$	$\uparrow 34.42 \pm 14.15s$	$\uparrow 32.42 \pm 10.29s$	$\uparrow 21.77 \pm 5.35s$	$\uparrow 23.99 \pm 7.51s$	$\uparrow 21.41 \pm 6.70s$
$\sigma = \pm$**2.0** [cm]	$\uparrow 22.49 \pm 4.72s$	$\uparrow 16.85 \pm 8.89s$	$\uparrow 37.16 \pm 16.10s$	$\uparrow 36.29 \pm 13.20s$	$\uparrow 21.79 \pm 5.45s$	$\uparrow 25.70 \pm 10.08s$	$\uparrow 22.21 \pm 7.41s$
$\sigma = \pm$**2.5** [cm]	$\uparrow 22.86 \pm 5.42s$	$\uparrow 18.22 \pm 10.25s$	$\uparrow 40.41 \pm 17.95s$	$\uparrow 40.98 \pm 15.56s$	$\uparrow 22.32 \pm 6.05s$	$\uparrow 27.19 \pm 11.92s$	$\uparrow 23.47 \pm 9.22s$

Table A.8: **Duration of the full sequential tasks by ROMAN and all included experts within the sequence of actions.** The table reports the average duration, in seconds as well as the standard deviation of the full sequential scenario cases. Results were obtained amongst 1,000 trials per cell at noise levels ranging from $\sigma = \pm 0.0$ [cm] to $\sigma = \pm 2.5$ [cm] with 0.5 [cm] increments. Symbols of $\downarrow$ and $\uparrow$ denote a decrease and increase in time difference compared to the previous noise levels respectively.

The Duration of the Full Sequential Tasks of ROMAN							
	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**	**Scenario 7**
	One Expert	*Two Expert*	*Three Expert*	*Four Expert*	*Five Expert*	*Six Expert*	*Seven Expert*
Noise Levels	*Push-Button*	*+Push*	*+Pick & Insert*	*+Pick & Place*	*+Rotate Open*	*+Pick & Drop*	*+Pull-Open*
$\sigma = \pm$**0.0** [cm]	$23.44 \pm 8.78s$	$47.63 \pm 16.34s$	$93.97 \pm 36.55s$	$113.74 \pm 29.94s$	$178.18 \pm 72.36s$	$165.00 \pm 46.29s$	$210.29 \pm 64.49s$
$\sigma = \pm$**0.5** [cm]	$^{\downarrow}22.80 \pm 8.37s$	$^{\downarrow}45.34 \pm 15.12s$	$^{\uparrow}94.07 \pm 38.41s$	$^{\downarrow}110.94 \pm 27.45s$	$^{\downarrow}168.76 \pm 66.81s$	$^{\downarrow}163.41 \pm 42.59s$	$^{\downarrow}194.88 \pm 50.44s$
$\sigma = \pm$**1.0** [cm]	$^{\downarrow}21.76 \pm 7.09s$	$^{\downarrow}43.36 \pm 12.23s$	$^{\uparrow}95.89 \pm 40.36s$	$^{\uparrow}111.55 \pm 31.06s$	$^{\downarrow}168.20 \pm 69.28s$	$^{\uparrow}168.79 \pm 50.85s$	$^{\downarrow}192.29 \pm 50.12s$
$\sigma = \pm$**1.5** [cm]	$^{\downarrow}21.46 \pm 6.89s$	$^{\downarrow}42.56 \pm 12.39s$	$^{\uparrow}104.76 \pm 44.33s$	$^{\uparrow}123.30 \pm 44.86s$	$^{\uparrow}176.17 \pm 71.14s$	$^{\uparrow}183.79 \pm 69.00s$	$^{\uparrow}211.75 \pm 79.52s$
$\sigma = \pm$**2.0** [cm]	$^{\uparrow}21.59 \pm 7.37s$	$^{\downarrow}41.75 \pm 12.64s$	$^{\downarrow}103.96 \pm 44.37s$	$^{\uparrow}125.12 \pm 47.27s$	$^{\downarrow}171.14 \pm 69.50s$	$^{\uparrow}194.38 \pm 74.57s$	$^{\uparrow}220.25 \pm 86.50s$
$\sigma = \pm$**2.5** [cm]	$^{\downarrow}21.36 \pm 6.73s$	$^{\downarrow}41.40 \pm 13.65s$	$^{\uparrow}104.22 \pm 44.55s$	$^{\uparrow}136.62 \pm 54.38s$	$^{\uparrow}181.10 \pm 72.83s$	$^{\uparrow}213.84 \pm 85.52s$	$^{\uparrow}236.69 \pm 97.32s$

Table A.9: **The architectural overview and details of ROMAN in its preliminary 2D setting composed of five experts.** This table reports the states and actions of each individual expert and the MN's of the ROMAN architecture, including their individual dimensions.

Preliminary ROMAN stage in 2D Consisting of Five Experts					
Network Architecture, Characteristics and Demonstration Settings					
MN	**Experts NNs**				
	Push	Lift	Pick & Place	Pick & Drop	Pull
State Space (Vector Size)					
Total: 29	***Total: 11***	***Total: 9***	***Total: 11***	***Total: 11***	***Total: 9***
Agent Position (2)	Agent Position (2)	Agent Position (2)	Agent Position (2)	Agent Position (2)	Agent Position (2)
Agent Velocity (2)	Agent Velocity (2)	Agent Velocity (2)	Agent Velocity (2)	Agent Velocity (2)	Agent Velocity (2)
Gripper Force (2)	Gripper Force (2)	Gripper Force (2)	Gripper Force (2)	Gripper Force (2)	Gripper Force (2)
Gripper State (1)	Gripper State (1)	Gripper State (1)	Gripper State (1)	Gripper State (1)	Gripper State (1)
† Full Environment (14)	Rack Position (2)	Gate Position (2)	Rack Position (2)	Box Position (2)	Drawer Position (2)
	Rack Target Location (2)		Vial Position (2)	Unbox Target Location (2)	
Action Space (Vector Size)					
Total: 5	***Total: 3***	***Total: 3***	***Total: 3***	***Total: 3***	***Total: 3***
Agent Weights (5)	Agent Velocity (2)	Agent Velocity (2)	Agent Velocity (2)	Agent Velocity (2)	Agent Velocity (2)
	Gripper State (1)	Gripper State (1)	Gripper State (1)	Gripper State (1)	Gripper State (1)
† Full Environment (14): Combined proprioceptive state space of all experts.					

Table A.10: **The architecture overview of all incorporated NNs in the hierarchical formation of ROMAN.** The table reports the NN architectures of the incorporated expert NNs and the MN's in the ROMAN architecture. In the employed hybrid learning procedure, the NNs are represented, in the context of GAIL, as the generators. The respective discriminator and its characteristics within GAIL are also detailed. Furthermore, the MLP for each is listed in the table. In accordance with the detailed expansion of the hybrid learning procedure in the main body of the thesis and in particular the modification to GAIL's discriminator of only differentiating states (s_t) and not the addition of actions (a_t), the number of inputs for the generator and discriminator are identical. **Notation †:** Expert dependent. It is worth noting, as per the main body of the thesis, that the state space of the expert NNs is decided on their individual sub-task goal, allowing these NNs, in turn, to solely focus on their respective end goal and omit non-relevant exteroceptive states. Due to the diverse nature of these expert NNs concerned with different types of manipulation skills, this naturally drove the need for different number of demonstrations and training times for each expert skill, which was primarily influenced by the higher-level task complexity those were concerned with. **Notation ‡:**: The MN observed the combined state space of the incorporated experts in the hierarchical architecture and ultimately the full environment state as to oversee and supervise the long-horizon sequential task.

ROMAN's NN Architectures			
Expert Networks		**Manipulation Network**	
State Space (Vector Size)		**State Space (Vector Size)**	
Agent Position	(3)	Agent Position	(3)
Agent Velocity	(3)	Agent Velocity	(3)
Gripper Force	(2)	Gripper Force	(2)
Environment State	(3-6)†	Full Environment State	(21)‡
Action Space (Vector Size)		**Action Space (Vector Size)**	
Agent Velocity	(3)	Agent Weights	(7)
Gripper State	(1)		
Demonstration Settings		**Demonstration Settings**	
Number of Demos	N = 20	Number of Demos	N = 42
Demo Time	$t \approx$ 6-12min†	Demo Time	$t \approx$ 42min
Generator (Expert Networks)		**Generator (Manipulation Network)**	
Number of Inputs	11 to 14†	Number of Inputs	29
Number of Outputs	4	Number of Outputs	7
Number of Hidden Layers	3	Number of Hidden Layers	3
Hidden Units Per Layer	128 - 256†	Hidden Units Per Layer	256
Discriminator (GAIL)		**Discriminator (GAIL)**	
Number of Inputs	11 to 14†	Number of Inputs	29
Number of Outputs	1	Number of Outputs	1
Number of Hidden Layers	2	Number of Hidden Layers	2
Hidden Units Per Layer	128	Hidden Units Per Layer	128

Table A.11: **The hyperparameters used for all experts in the hierarchical architecture of ROMAN, including the gating network (MN).** This table reports the hyperparameter values for all experts and that of the gating network, as well as the NN structure (layers) used for each within the ROMAN architecture. The hyperparameters were identical for all experts and differed for the MN. As described in the main body of the thesis, BC was used to warm-start the policy prior to using PPO and thereafter a GAIL reward was used for the PPO policy, in the form of an intrinsic reward $r^i = -log(1 - D(s_t))$.

ROMAN's Hyperparameters Settings (PPO)			
Expert Networks		**Manipulation Network**	
Minibatch Range	1,024	Minibatch Range	1,024
GAE Parameter Lambda Range	0.95	GAE Parameter Lambda Range	0.95
Entropy Coefficient Range	5.0e-3	Entropy Coefficient Range	5.0e-3
Horizon Range	1,000	Horizon Range	1,000
Number of Epochs	3	Number of Epochs	3
Clipping Parameter Epsilon	0.2	Clipping Parameter Epsilon	0.2
Discount Factor Gamma Range	0.99	Discount Factor Gamma Range	0.99
Learning Rate	3.0e-4	Learning Rate	3.0e-4
Replay Buffer Observation Size	10,240	Replay Buffer Observation Size	10,240
ROMAN's Hyperparameters Settings (GAIL - Discriminator)			
Expert Networks		**Manipulation Network**	
Discount Factor Gamma Range	0.99	Discount Factor Gamma Range	0.99
Learning Rate	3.0e-4	Learning Rate	3.0e-4
ROMAN's Hyperparameters Settings (Behavioural Cloning)			
Expert Networks		**Manipulation Network**	
Batch Size	1,024	Batch Size	1,024
Learning Rate	3.0e-4	Learning Rate	3.0e-4

Table A.12: **Architecture settings for the monolithic NNs for both the 2D and 3D cases.** The table reports the architectural settings of the monolithic NNs for both 2D and 3D cases used as part of the baseline evaluation against the hierarchical architecture of ROMAN. More specifically, the state and action spaces are detailed, including the demonstrations provided to each monolithic NN. The state space of the monolithic NNs was of identical size to that of the MN. In contrast to ROMAN's MN which controlled the weights of expert NNs, the action space of the monolithic NNs directly controlled the velocity of the end-effector and the opening/closing of the gripper. The same hybrid learning approach was employed for both ROMAN's NNs and the monolithic NNs to further retain consistency.

Monolithic NN Architectures			
Single NN in 2D		**Single NN in 3D**	
State Space (Vector Size) †		**State Space (Vector Size) †**	
Agent Position	(2)	Agent Position	(3)
Agent Velocity	(2)	Agent Velocity	(3)
Gripper Force	(2)	Gripper Force	(2)
Gripper State	(1)	Button Position	(3)
Rack Position	(2)	Rack Position	(3)
Gate Position	(2)	Conveyor Position	(3)
Vial Position	(2)	Vial Position	(3)
Box Position	(2)	Box Position	(3)
Unbox Target Location	(2)	Drawer Position	(3)
Drawer Position	(2)	Cabinet Position	(3)
Rack Target Location	(2)		
Action Space (Vector Size) ‡		**Action Space (Vector Size) ‡**	
Agent Velocity	(2)	Agent Velocity	(3)
Gripper State	(1)	Gripper State	(1)
Demonstration Settings		**Demonstration Settings**	
Number of Demos	N = 100	Number of Demos	N = 140
Demo Time	$t \approx 64$min	Demo Time	$t \approx 132$min
Generator (Single NN in 2D)		**Generator (Single NN in 3D)**	
Number of Inputs	21	Number of Inputs	29
Number of Outputs	3	Number of Outputs	4
Number of Hidden Layers	3	Number of Hidden Layers	3
Hidden Units Per Layer	128	Hidden Units Per Layer	256
Discriminator (GAIL)		**Discriminator (GAIL)**	
Number of Inputs	21	Number of Inputs	29
Number of Outputs	1	Number of Outputs	1
Number of Hidden Layers	2	Number of Hidden Layers	2
Hidden Units Per Layer	64	Hidden Units Per Layer	128

† Identical to ROMAN's MN state space for 2D/3D accordingly.

‡ Identical to ROMAN's experts action space for 2D/3D accordingly.

Table A.13: **Hyperparameters for the monolithic NNs for both the 2D and 3D cases.** The table reports the hyperparameter values for the monolithic NNs used as a baseline evaluation against ROMAN. To retain consistency during the baseline evaluation of monolithic NNs vs the hierarchical formation of ROMAN, identical hyperparameter values were used for the monolithic NNs to that of ROMAN's MN for the 2D and 3D cases. An identical hybrid learning procedure was used for the monolithic NNs as with ROMAN's hierarchical formation.

Monolithic NN Hyperparameters Settings (PPO)			
Single NN (2D)		**Single NN (3D)**	
Minibatch Range	1,024	Minibatch Range	1,024
GAE Parameter Lambda Range	0.95	GAE Parameter Lambda Range	0.95
Entropy Coefficient Range	5.0e-3	Entropy Coefficient Range	5.0e-3
Horizon Range	1,000	Horizon Range	1,000
Number of Epochs	3	Number of Epochs	3
Clipping Parameter Epsilon	0.2	Clipping Parameter Epsilon	0.2
Discount Factor Gamma Range	0.99	Discount Factor Gamma Range	0.99
Learning Rate	3.0e-4	Learning Rate	3.0e-4
Replay Buffer Observation Size	10,240	Replay Buffer Observation Size	10,240
Monolithic NN Hyperparameters Settings (GAIL - Discriminator)			
Single NN (2D)		**Single NN (3D)**	
Discount Factor Gamma Range	0.99	Discount Factor Gamma Range	0.99
Learning Rate	3.0e-4	Learning Rate	3.0e-4
Monolithic NN Hyperparameters Settings (Behavioural Cloning)			
Single NN (2D)		**Single NN (3D)**	
Batch Size	512	Batch Size	1024
Learning Rate	3.0e-4	Learning Rate	3.0e-4

Appendix B

Additional Materials – IGE-LLMs

In this appendix, additional technical details and elaboration of the IGE-LLMs framework are provided, which were not deemed critical to the main understanding of the framework.

B.1 LLM Prompts

This section presents the prompts that were used as inputs to the LLM. The prompts that were provided to the LLM both as a direct policy driver (by employing the argmax on the generated dictionary) and for the generation of the intrinsic reward dictionaries via the IGE-LLMs framework are detailed in Figure B.1 and Figure B.2, corresponding to the DeepSea and ROMAN environments respectively.

IGE-LLMs (DeepSea Environment) - General Prompt

Textual Description of the Goals and Constraints as well as States and Actions

(Initial and Recurring Prompts are Merged due to the Context Simplicity of the Environment)

You are tasked with evaluating the agent navigating in the DeepSea environment and more specifically its actions.

The agent starts in the top left and has to reach the goal in the bottom right location. Grid size is *NxN*.

The current agent position is:
X,Y

Possible actions:
1. Down (x, y+1)
2. Diagonal Down Right (x+1, y+1)

Evaluate each action from 0 to 1. Explain the ratings and be very brief with your response.

Figure B.1: **Textual description provided to the LLM for the DeepSea environment.**

IGE-LLMs (ROMAN Environment) - Initial Prompt

Textual Description of the Goals and Constraints
(Specific Prompt - Task Dependent)

You are evaluating the actions of a gating network.

State of the environment:
In the laboratory room, the workspace contains a cabinet, a drawer, a box, a cover, a rack, a vial, a table, a button, and a conveyor belt. The drawer is closed, and the box is inside the drawer. The cover is on top of the box, and the vial is inside the box. The cabinet door is closed, and the rack is inside the cabinet, which is located directly underneath the table. The table is right next to the conveyor belt at the same height. The button is located near the conveyor belt and is used to activate it. The conveyor belt is by default inactive unless the button is pressed.

Constraints:
There are obstacles over the conveyor belt and the cabinet clearance is limited. The vial cannot be inserted in the rack unless the rack is on the table.

Unless otherwise stated, assume the environment is as described above. If certain objects are not visible, assume these are occluded.

The possible actions by the gating network are the following:
1. Pull (Pulls Open the Drawer)
2. Pick and Drop (Unbox / Removes the Cover of the Box)
3. Rotate Open (Rotates Open the Cabinet Door)
4. Pick and Place (Places the Rack)
5. Pick and Insert (Inserts the Vial)
6. Push (Pushes the Rack with the Vial)
7. Push Button (Activates the Conveyor Belt)

End-Goal:
The end goal is for the gating network to automate a sequence, whereby the rack will be securely housing the vial, which will be moved to another part of the production line that is connected via the conveyor belt.

Your inputs:
1. The state of the environment.
2. The seven possible actions that can be performed by the gating network.

Your outputs:
Based on the state of the environment which will be provided to you after this prompt, you are tasked with evaluating the seven possible actions performed by the gating network. Your output should be a float value from 0 to 1, with 0 being an incorrect or improper action while 1 indicating a correct action.

Ready to commence and be provided the states of the environment?

IGE-LLMs (ROMAN Environment) - Recurring Prompt

Textual Description of States and Actions
(General Prompt - Task Independent)

The current environment state is:
Conveyor Button: Not pressed, inactive.
Cabinet: Closed.
Rack: Inside cabinet.
Drawer: Closed.
Box: Closed, inside drawer.
Vial: Within the drawer, inside the box.

Given the Following Actions:
1. Pull (Pulls Open the Drawer)
2. Pick and Drop (Unbox / Removes the Cover of the Box)
3. Rotate Open (Rotates Open the Cabinet Door)
4. Pick and Place (Places the Rack)
5. Pick and Insert (Inserts the Vial)
6. Push (Pushes the Rack with the Vial)
7. Push Button (Activates the Conveyor Belt)

Some of the actions may be applicable and some not (either redundant or incorrect). For each of the seven actions, rate them from 0 to 1. In addition to each of the ratings, provide a very brief one sentence description why.

Figure B.2: **Textual descriptions provided to the LLM for the ROMAN environment.**

Bibliography

A Ghazanfar, A. and E Schroeder, C. (2006). Is neocortex essentially multisensory? *Trends in cognitive sciences*, 10:278–85.

Abbeel, P. and Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the Twenty-First International Conference on Machine Learning*, ICML '04, page 1, New York, NY, USA. Association for Computing Machinery.

Ahn, M., Brohan, A., Brown, N., Chebotar, Y., Cortes, O., David, B., Finn, C., Fu, C., Gopalakrishnan, K., Hausman, K., Herzog, A., Ho, D., Hsu, J., Ibarz, J., Ichter, B., Irpan, A., Jang, E., Ruano, R. J., Jeffrey, K., Jesmonth, S., Joshi, N. J., Julian, R., Kalashnikov, D., Kuang, Y., Lee, K.-H., Levine, S., Lu, Y., Luu, L., Parada, C., Pastor, P., Quiambao, J., Rao, K., Rettinghouse, J., Reyes, D., Sermanet, P., Sievers, N., Tan, C., Toshev, A., Vanhoucke, V., Xia, F., Xiao, T., Xu, P., Xu, S., Yan, M., and Zeng, A. (2022). Do as i can, not as i say: Grounding language in robotic affordances.

Aleotti, J., Bottazzi, S., and Reggiani, M. (2002). A multimodal user interface for remote object exploration in teleoperation systems.

Amedi, A., Malach, R., Hendler, T., Peled, S., and Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience*, 4:324–330.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané, D. (2016). Concrete problems in ai safety.

Andrychowicz, O. M., Baker, B., Chociej, M., Józefowicz, R., McGrew, B., Pachocki, J., Petron, A., Plappert, M., Powell, G., Ray, A., Schneider, J., Sidor, S., Tobin, J., Welinder, P., Weng, L., and Zaremba, W. (2020). Learning dexterous in-hand manipulation. *The International Journal of Robotics Research*, 39(1):3–20.

Antotsiou, D., Ciliberto, C., and Kim, T. (2022). Modular adaptive policy selection for

multi- task imitation learning through task division. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2459–2465.

Argelaguet, F. and Andujar, C. (2013). A survey of 3d object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136.

Ashe, J., Lungu, O. V., Basford, A. T., and Lu, X. (2006). Cortical control of motor sequences. *Current Opinion in Neurobiology*, 16(2):213–221.

Azmandian, M., Hancock, M., Benko, H., Ofek, E., and Wilson, A. D. (2016). Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 1968–1979, New York, NY, USA. ACM.

B F van Erp, J. and Padmos, P. (2004). Image parameters for driving with indirect viewing systems. *Ergonomics*, 46:1471–99.

Barfield, W. (1995). *Virtual environments and advanced interface design*. Oxford University Press.

Barrera Machuca, M. D. and Stuerzlinger, W. (2019). The effect of stereo display deficiencies on virtual hand pointing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, New York, NY, USA. Association for Computing Machinery.

Batmaz, A. U., Machuca, M. D. B., Pham, D. M., and Stuerzlinger, W. (2019). Do head-mounted display stereo deficiencies affect 3d pointing tasks in ar and vr? In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 585–592, Osaka, Japan. IEEE.

Behbahani, F., Shiarlis, K., Chen, X., Kurin, V., Kasewa, S., Stirbu, C., Gomes, J., Paul, S., Oliehoek, F. A., Messias, J., and Whiteson, S. (2019). Learning from demonstration in the wild. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 775–781.

Bellemare, M. G., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., and Munos, R. (2016). Unifying count-based exploration and intrinsic motivation.

Berger, C., Gonzalez-Franco, M., Ofek, E., and Hinckley, K. (2018). The uncanny valley of haptics. *Science Robotics*, 3:eaar7010.

Billard, A. and Kragic, D. (2019). Trends and challenges in robot manipulation. *Science*, 364(6446).

Blackman, H. S., Nelson, W. R., and Hahn, H. A. (1992). Measurement of human performance in complex task environments. *Human Performance*, 5(4):329–351.

Bolopion, A. and Régnier, S. (2013). A review of haptic feedback teleoperation systems for micromanipulation and microassembly. *IEEE Transactions on Automation Science and Engineering*, 10(3):496–502.

Brickler, D., Babu, S. V., Bertrand, J., and Bhargava, A. (2018). Towards evaluating the effects of stereoscopic viewing and haptic interaction on perception-action coordination. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1–516.

Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

Brooks, J., Lodge, R., and White, D. (2017). Comparison of a head-mounted display and flat screen display during a micro-uav target detection task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1):1514–1518.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Bunt, H., Beun, R.-J., and Borghuis, T. (1998). *Multimodal human-computer communication: systems, techniques, and experiments*, volume 1374. Springer Science & Business Media, Berlin-Heidelberg, Germany.

Burda, Y., Edwards, H., Storkey, A., and Klimov, O. (2019). Exploration by random network distillation. In *International Conference on Learning Representations*.

Burdea, G., Richard, P., and Coiffet, P. (1996). Multimodal virtual reality: Input output devices, system integration, and human factors. *International Journal of Human Computer Interaction*, 8(1):5–24.

Burke, J. L., Prewett, M. S., Gray, A. A., Yang, L., Stilson, F. R. B., Coovert, M. D., Elliot, L. R., and Redden, E. (2006). Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A meta-analysis. In *Proceedings of the 8th International Conference on Multimodal Interfaces*, ICMI '06, pages 108–117, New York, NY, USA. ACM.

Burno, R. A., Wu, B., Doherty, R., Colett, H., and Elnaggar, R. (2015). Applying fitts' law to gesture based computer interactions. *Procedia Manufacturing*, 3:4342 – 4349. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015.

Burns, E., Razzaque, S., Whitton, M. C., and Brooks, F. P. (2007). Macbeth: The avatar which i see before me and its movement toward my hand. In *2007 IEEE Virtual Reality Conference*, pages 295–296.

Burstyn, J., Carrascal, J. P., and Vertegaal, R. (2016). Fitts' law and the effects of input mapping and stiffness on flexible display interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 3649–3658, New York, NY, USA. Association for Computing Machinery.

Carta, T., Romac, C., Wolf, T., Lamprier, S., Sigaud, O., and Oudeyer, P.-Y. (2023). Grounding large language models in interactive environments with online reinforcement learning.

Cha, Y. and Myung, R. (2013). Extended fitts' law for 3d pointing tasks using 3d target arrangements. *International Journal of Industrial Ergonomics*, 43(4):350 – 355.

Chebotar, Y., Handa, A., Makoviychuk, V., Macklin, M., Issac, J., Ratliff, N., and Fox, D. (2019). Closing the sim-to-real loop: Adapting simulation randomization with real world experience. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8973–8979.

Chen, J. Y. C., Haas, E. C., and Barnes, M. J. (2007). Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(6):1231–1245.

Chentanez, N., Barto, A., and Singh, S. (2004). Intrinsically motivated reinforcement learning. In Saul, L., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems*, volume 17. MIT Press.

Choi, I., Hawkes, E. W., Christensen, D. L., Ploch, C. J., and Follmer, S. (2016). Wolverine: A wearable haptic interface for grasping in virtual reality. *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Cholewiak, R. W. and Collins, A. A. (2000). The generation of vibrotactile patterns on a linear array: Influences of body site, time, and presentation mode. *Perception & Psychophysics*, 62(6):1220–1235.

Codevilla, F., Santana, E., Lopez, A., and Gaidon, A. (2019). Exploring the limitations of behavior cloning for autonomous driving. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9328–9337.

Davchev, T., Sushkov, O. O., Regli, J.-B., Schaal, S., Aytar, Y., Wulfmeier, M., and Scholz, J. (2022). Wish you were here: Hindsight goal selection for long-horizon dexterous manipulation. In *International Conference on Learning Representations*.

de Grosbois, J., Heath, M., and Tremblay, L. (2015). Augmented feedback influences upper limb reaching movement times but does not explain violations of fitts' law. *Frontiers in Psychology*, 6:800.

DeJong, B. P., Colgate, J. E., and Peshkin, M. A. (2004). Improving teleoperation: reducing mental rotations and translations. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, volume 4, pages 3708–3714 Vol.4.

Deng, C.-L., Geng, P., Hu, Y.-F., and Kuai, S.-G. (2019). Beyond fitts's law: A three-phase model predicts movement time to position an object in an immersive 3d virtual environment. *Human Factors*, 61(6):879–894. PMID: 30912987.

Ding, Y., Florensa, C., Abbeel, P., and Phielipp, M. (2019). Goal-conditioned imitation learning. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Drewes, H. (2010). Only one fitts' law formula please! In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, page 2813–2822, New York, NY, USA. Association for Computing Machinery.

Du, Y., Watkins, O., Wang, Z., Colas, C., Darrell, T., Abbeel, P., Gupta, A., and Andreas, J. (2023). Guiding pretraining in reinforcement learning with large language models.

Durgin, F. H., Proffitt, D. R., Olson, T. J., and Reinke, K. S. (1995). Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3):679.

Erni, T. and Dietz, V. (2001). Obstacle avoidance during human walking: learning rate and cross-modal transfer. *The Journal of physiology*, 534(1):303–312.

Fani, S., Ciotti, S., Catalano, M. G., Grioli, G., Tognetti, A., Valenza, G., Ajoudani, A., and Bianchi, M. (2018). Simplifying telerobotics: Wearability and teleimpedance improves human-robot interactions in teleoperation. *IEEE Robotics Automation Magazine*, 25(1):77–88.

Ferrari, C. and Canny, J. (1992). Planning optimal grasps. In *Proceedings 1992 IEEE International Conference on Robotics and Automation*, pages 2290–2295 vol.3.

Ferreira, A. and Mavroidis, C. (2006). Virtual reality and haptics for nanorobotics. *IEEE Robotics Automation Magazine*, 13(3):78–92.

Finn, C., Levine, S., and Abbeel, P. (2016). Guided cost learning: Deep inverse optimal control via policy optimization. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, page 49–58.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381.

Fitts, P. M. and Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of experimental psychology*, 67(2):103.

Flanagan, J. R., Bowman, M. C., and Johansson, R. S. (2006). Control strategies in object manipulation tasks. *Current Opinion in Neurobiology*, 16(6):650–659.

Flet-Berliac, Y., Ferret, J., Pietquin, O., Preux, P., and Geist, M. (2021). Adversarially guided actor-critic. In *International Conference on Learning Representations*.

Forlines, C. and Balakrishnan, R. (2008). Evaluating tactile feedback and direct vs. indirect stylus input in pointing and crossing selection tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, page 1563–1572, New York, NY, USA. Association for Computing Machinery.

Fox, R., Berenstein, R., Stoica, I., and Goldberg, K. (2019). Multi-task hierarchical

imitation learning for home automation. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, pages 1–8.

Fox, R., Shin, R., Krishnan, S., Goldberg, K., Song, D., and Stoica, I. (2018). Parametrized hierarchical procedures for neural programming. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*.

Frans, K., Ho, J., Chen, X., Abbeel, P., and Schulman, J. (2018). Meta learning shared hierarchies. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*.

Frid, E., Moll, J., Bresin, R., and Sallnäs Pysander, E.-L. (2018). Haptic feedback combined with movement sonification using a friction sound improves task performance in a virtual throwing task. *Journal on Multimodal User Interfaces*.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701.

Fukuda, T., Tanie, K., Mitsuoka, T., Hashimoto, H., and Harashima, F. (1988). Dual mode control method of micro-manipulator with visual feedback. *IFAC Proceedings Volumes*, 21(16):377 – 382. 2nd IFAC Symposium on Robot Control 1988 (SYROCO '88), Karlsruhe, FRG, 5-7 October.

Gawron, V. J., Cipriano, L., Fleishman, E., Hegge, F., Lehman, E., Meister, D., and Reising, J. (1990). Guide for human performance measurements. *Proceedings of the Human Factors Society Annual Meeting*, 34(17):1238–1240.

Gemperle, F., Ota, N., and Siewiorek, D. (2001). Design of a wearable tactile display. In *Proceedings Fifth International Symposium on Wearable Computers*, pages 5–12.

Gillan, D. J., Holden, K., Adam, S., Rudisill, M., and Magee, L. (1992). How should fitts' law be applied to human-computer interaction? *Interacting with Computers*, 4(3):291 – 313.

Godse, A., Khadka, R., and Banic, A. (2019). Evaluation of visual perception manipulation in virtual reality training environments to improve golf performance. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1807–1812, Osaka, Japan. IEEE.

Goldberg, K. (2019). Robots and the return to collaborative intelligence. *Nature Machine Intelligence*, 1(1):2–4.

Gonzalez-Franco, M. and Peck, T. C. (2018). Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74.

Gori, J., Rioul, O., Guiard, Y., and Beaudouin-Lafon, M. (2018). The perils of confounding factors: How fitts' law experiments can lead to false conclusions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–10, New York, NY, USA. Association for Computing Machinery.

Gu, S., Lillicrap, T., Ghahramani, Z., Turner, R. E., Schölkopf, B., and Levine, S. (2017). Interpolated policy gradient: Merging on-policy and off-policy gradient estimation for deep reinforcement learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 3849–3858, Red Hook, NY, USA. Curran Associates Inc.

Gu, X., Zhang, Y., Sun, W., Bian, Y., Zhou, D., and Kristensson, P. O. (2016). Dexmo: An inexpensive and lightweight mechanical exoskeleton for motion capture and force feedback in vr. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 1991–1995, New York, NY, USA. Association for Computing Machinery.

Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1861–1870. PMLR.

Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. 52:139–183.

Heilig, M. L. (1992). El cine del futuro: The cinema of the future. *Presence: Teleoperators & Virtual Environments*, 1:279–294.

Hess, R. F., To, L., Zhou, J., Wang, G., and Cooperstock, J. R. (2015). Stereo vision: The haves and have-nots. *i-Perception*, 6(3):2041669515593028.

Ho, J. and Ermon, S. (2016). Generative adversarial imitation learning. In Lee, D.,

Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

Hoffmann, E. R. (1991). A comparison of hand and foot movement times. *Ergonomics*, 34(4):397–406. PMID: 1860460.

Hoffmann, E. R. (1995). Effective target tolerance in an inverted fitts task. *Ergonomics*, 38(4):828–836.

Hoffmann, E. R., Drury, C. G., and Romanowski, C. J. (2011). Performance in one-, two- and three-dimensional terminal aiming tasks. *Ergonomics*, 54(12):1175–1185. PMID: 22103725.

Hong, H. and Kang, S. H. (2015). Measurement of the lens accommodation in viewing stereoscopic displays. *Journal of the Society for Information Display*, 23(1):19–26.

James, T., Keith Humphrey, G., Gati, S., Servos, P., S Menon, R., and Goodale, M. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia*, 40:1706–14.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. (2008). Measuring and defining the experience of immersion in games. *Int. J. Hum.-Comput. Stud.*, 66(9):641–661.

Johnson, S., Rae, I., Mutlu, B., and Takayama, L. (2015). Can you see me now?: How field of view affects collaboration in robotic telepresence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2397–2406, New York, NY, USA. ACM.

Juliani, A., Berges, V.-P., Teng, E., Cohen, A., Harper, J., Elion, C., Goy, C., Gao, Y., Henry, H., Mattar, M., et al. (2018). Unity: A general platform for intelligent agents. *arXiv preprint arXiv:1809.02627*.

Kenyon, R. V. and Ellis, S. R. (2014). *Vision, perception, and object manipulation in virtual environments*. Springer, New York, USA.

Kerr, R. (1973). Movement time in an underwater environment. *Journal of Motor Behavior*, 5(3):175–178. PMID: 23961747.

Keshavarz, B., Stelzmann, D., Paillard, A., and Hecht, H. (2015). Visually induced

motion sickness can be alleviated by pleasant odors. *Experimental Brain Research*, 233(5):1353–1364.

Kilteni, K., Groten, R., and Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoper. Virtual Environ.*, 21(4):373–387.

Kim, M., Jeon, C., and Kim, J. (2017). A study on immersion and presence of a portable hand haptic system for immersive virtual reality. *Sensors*, 17(5):1141.

Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., and Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychological Science*, 9(4):293–298.

Koganti, N., Rahman H. A. G., A., Iwasawa, Y., Nakayama, K., and Matsuo, Y. (2018). Virtual reality as a user-friendly interface for learning from demonstrations. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, page 1–4, New York, NY, USA. Association for Computing Machinery.

Kondraske, G. V. (1994). An angular motion fitt's law for human performance modeling and prediction. In *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 307–308 vol.1, Baltimore, MD, USA. IEEE.

Kontarinis, D. A. and Howe, R. D. (1995). Tactile display of vibratory information in teleoperation and virtual environments. *Presence: Teleoper. Virtual Environ.*, 4(4):387–402.

Kulik, A., Kunert, A., and Froehlich, B. (2020). On motor performance in virtual 3d object manipulation. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):2041–2050.

Kumar, V. and Todorov, E. (2015). Mujoco haptix: A virtual reality system for hand manipulation. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pages 657–663, Seoul, South Korea. IEEE.

Kunert, A., Kulik, A., Huckauf, A., and Fröhlich, B. (2007). A comparison of tracking- and controller-based input for complex bimanual interaction in virtual environments. In *Proceedings of the 13th Eurographics Conference on Virtual Environments*, EGVE'07, page 43–52, Goslar, DEU. Eurographics Association.

Kwon, M., Xie, S. M., Bullard, K., and Sadigh, D. (2023). Reward design with language models.

Lampton, D. R., McDonald, D. P., Singer, M., and Bliss, J. P. (1995). Distance estimation in virtual environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 39(20):1268–1272.

Lathan, C. E. and Tracey, M. (2002). The effects of operator spatial perception and sensory feedback on human-robot teleoperation performance. *Presence: Teleoper. Virtual Environ.*, 11(4):368–377.

LaValle, S. M. (2019). *Birds-Eye View, In Virtual Reality*. Cambridge University Press.

Le, H. M., Jiang, N., Agarwal, A., Dudík, M., Yue, Y., and III, H. D. (2018). Hierarchical imitation and reinforcement learning. In Dy, J. G. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 2923–2932. PMLR.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lee, M. A., Zhu, Y., Srinivasan, K., Shah, P., Savarese, S., Fei-Fei, L., Garg, A., and Bohg, J. (2019). Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8943–8950.

Levine, S. and Abbeel, P. (2014). Learning neural network policies with guided policy search under unknown dynamics. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Light, C. M., Chappell, P. H., and Kyberd, P. J. (2002). Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: Normative data, reliability, and validity. *Archives of Physical Medicine and Rehabilitation*, 83(6):776–783.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. (2016). Continuous control with deep reinforcement learning. In Bengio, Y. and LeCun, Y., editors, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.

Lin, C. J. and Woldegiorgis, B. H. (2015). Interaction and visual performance in stereoscopic displays: A review. *Journal of the Society for Information Display*, 23(7):319–332.

Lipton, J. I., Fay, A. J., and Rus, D. (2018). Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing. *IEEE Robotics and Automation Letters*, 3(1):179–186.

Liu, L., van Liere, R., Nieuwenhuizen, C., and Martens, J. (2009). Comparing aimed movements in the real world and in virtual reality. In *2009 IEEE Virtual Reality Conference*, pages 219–222, Lafayette, LA, USA. IEEE.

Liu, Y., Gupta, A., Abbeel, P., and Levine, S. (2018). Imitation from observation: Learning to imitate behaviors from raw video via context translation. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1118–1125.

Lubos, P., Bruder, G., and Steinicke, F. (2014). Analysis of direct selection in head-mounted display environments. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 11–18, Minneapolis, MN, USA. IEEE.

Lupyan, G. and Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, 24(4):279–284.

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact.*, 7(1):91–139.

MacKenzie, I. S. and Isokoski, P. (2008). Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, page 1633–1636, New York, NY, USA. Association for Computing Machinery.

MacKenzie, I. S., Sellen, A., and Buxton, W. A. S. (1991). A comparison of input devices in element pointing and dragging tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, page 161–166, New York, NY, USA. Association for Computing Machinery.

Martins, H. and Ventura, R. (2009). Immersive 3-d teleoperation of a search and rescue robot using a head-mounted display. *2009 IEEE Conference on Emerging Technologies & Factory Automation*.

Marzari, L., Pore, A., Dall'Alba, D., Aragon-Camarasa, G., Farinelli, A., and Fiorini, P. (2021). Towards hierarchical task decomposition using deep reinforcement learning for pick and place subtasks. In *2021 20th International Conference on Advanced Robotics (ICAR)*, pages 640–645.

Massimino, M. J. and Sheridan, T. B. (1993). Sensory substitution for force feedback in teleoperation. *Presence: Teleoperators and Virtual Environments*, 2(4):344–352.

McGee, M., Amento, B., Brooks, P., and Harley, H. (1997). Fitts and vr: Evaluating display and input devices with fitts' law. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 41(2):1259–1262.

McGlynn, S. A. and Rogers, W. A. (2017). Considerations for presence in teleoperation. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17, page 203–204, New York, NY, USA. Association for Computing Machinery.

McIntire, J. P., Havig, P. R., and Geiselman, E. E. (2014). Stereoscopic 3d displays and human performance: A comprehensive review. *Displays*, 35(1):18 – 26.

Menchaca-Brandan, M. A., Liu, A. M., Oman, C. M., and Natapoff, A. (2007). Influence of perspective-taking and mental rotation abilities in space teleoperation. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, HRI '07, page 271–278, New York, NY, USA. Association for Computing Machinery.

Merel, J., Ahuja, A., Pham, V., Tunyasuvunakool, S., Liu, S., Tirumala, D., Heess, N., and Wayne, G. (2019a). Hierarchical visuomotor control of humanoids. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*.

Merel, J., Botvinick, M., and Wayne, G. (2019b). Hierarchical motor control in mammals and machines. *Nature Communications*, 10(1):5489.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Harley, T., Lillicrap, T. P., Silver, D., and Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, page 1928–1937.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A.,

Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100.

Murata, A. and Iwase, H. (2001). Extending fitts' law to a three-dimensional pointing task. *Human movement science*, 20(6):791–805.

Mülling, K., Kober, J., Kroemer, O., and Peters, J. (2013). Learning to select and generalize striking movements in robot table tennis. *The International Journal of Robotics Research*, 32(3):263–279.

Nagai, Y., Kimura, S., Tsuchiya, S., and Iida, T. (1999). Evaluation of audio feedback system for engineering test satellite vii. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.99CH37028)*, volume 5, pages 1135–1140 vol.5.

Narumi, T., Nishizaka, S., Kajinami, T., Tanikawa, T., and Hirose, M. (2011). Augmented reality flavors: Gustatory display based on edible marker and cross-modal interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 93–102, New York, NY, USA. ACM.

O'Hara, J. M. (1987). Telerobotic control of a dextrous manipulator using master and six-dof hand controllers for space assembly and servicing tasks. *Proceedings of the Human Factors Society Annual Meeting*, 31(7):791–795.

Ortenzi, V., Controzzi, M., Cini, F., Leitner, J., Bianchi, M., Roa, M. A., and Corke, P. (2019). Robotic manipulation and the role of the task in the metric of success. *Nature Machine Intelligence*, 1(8):340–346.

Osband, I., Doron, Y., Hessel, M., Aslanides, J., Sezener, E., Saraiva, A., McKinney, K., Lattimore, T., Szepesvari, C., Singh, S., Roy, B. V., Sutton, R., Silver, D., and Hasselt, H. V. (2020). Behaviour suite for reinforcement learning. In *International Conference on Learning Representations*.

Ostrovski, G., Bellemare, M. G., van den Oord, A., and Munos, R. (2017). Count-based exploration with neural density models. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 2721–2730.

Oudeyer, P.-Y., Kaplan, F., and Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(2):265–286.

Pamungkas, D. S. and Ward, K. (2013). Tele-operation of a robot arm with electro tactile feedback. In *2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pages 704–709.

Pastor, P., Hoffmann, H., Asfour, T., and Schaal, S. (2009). Learning and generalization of motor skills by learning from demonstration. In *2009 IEEE International Conference on Robotics and Automation*, pages 763–768.

Pathak, D., Agrawal, P., Efros, A. A., and Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 2778–2787.

Peng, X. B., Chang, M., Zhang, G., Abbeel, P., and Levine, S. (2019). MCP: learning composable hierarchical control with multiplicative compositional policies. In Wallach, H. M., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. B., and Garnett, R., editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 3681–3692.

Poletti, C., Sleimen-Malkoun, R., Decker, L. M., Retornaz, F., Lemaire, P., and Temprado, J.-J. (2017). Strategic variations in fitts' task: comparison of healthy older adults and cognitively impaired patients. *Frontiers in aging neuroscience*, 8:334.

Popescu, G. V., Burdea, G. C., and Trefftz, H. (2002). *Multimodal interaction modeling.*, pages 435–454. Human factors and ergonomics. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US.

Poupyrev, I. and Ichikawa, T. (1999). Manipulating objects in virtual worlds: Categorization and empirical evaluation of interaction techniques. *Journal of Visual Languages & Computing*, 10(1):19 – 35.

Pukelsheim, F. (1994). The three sigma rule. *The American Statistician*, 48(2):88–91.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., McCullough, K. E., and Ansari, R. (2002). Multimodal human discourse: Gesture and speech. *ACM Trans. Comput.-Hum. Interact.*, 9(3):171–193.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. (2021). Learning transferable visual models from natural language supervision.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Rae, I., Venolia, G., Tang, J. C., and Molnar, D. (2015). A framework for understanding and designing telepresence. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 1552–1566, New York, NY, USA. ACM.

Rajeswaran, A., Kumar, V., Gupta, A., Vezzani, G., Schulman, J., Todorov, E., and Levine, S. (2018). Learning complex dexterous manipulation with deep reinforcement learning and demonstrations. In *Proceedings of Robotics: Science and Systems*, Pittsburgh, Pennsylvania.

Ranscombe, C., Rodda, J., and Johnson, M. (2019). Visualising user experiences: Analysing embodiment of ux in autonomous vehicle concepts. *Proceedings of the Design Society: International Conference on Engineering Design*, 1(1):4039–4048.

Ratliff, N., Bagnell, J. A., and Srinivasa, S. S. (2007). Imitation learning for locomotion and manipulation. In *2007 7th IEEE-RAS International Conference on Humanoid Robots*, pages 392–397.

Reddy, S., Dragan, A. D., and Levine, S. (2020). SQIL: imitation learning via reinforcement learning with sparse rewards. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*.

Renner, R. S., Velichkovsky, B. M., and Helmert, J. R. (2013). The perception of egocentric distances in virtual environments - a review. *ACM Comput. Surv.*, 46(2).

Richard, P., Burdea, G., Gomez, D., and Coiffet, P. (1994). A comparison of haptic, visual and auditive force feedback for deformable virtual objects. In *Proceedings of the Internation Conference on Automation Technology (ICAT)*, volume 49, page 62.

Rochlis, J. L. and Newman, D. J. (2000). A tactile display for international space station (iss) extravehicular activity (eva). *Aviation, space, and environmental medicine*, 71(6):571–578.

Rock, I. and Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, 143(3606):594–596.

Rosen, E., Whitney, D., Phillips, E., Chien, G., Tompkin, J., Konidaris, G., and Tellex, S. (2019). Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays. *The International Journal of Robotics Research*, 0(0):0278364919842925.

Rosenberg, L. B. (1993). The effect of interocular distance upon operator performance using stereoscopic displays to perform virtual depth tasks. In *Proceedings of IEEE Virtual Reality Annual International Symposium*, pages 27–32.

Ross, S., Gordon, G., and Bagnell, D. (2011). A reduction of imitation learning and structured prediction to no-regret online learning. In Gordon, G., Dunson, D., and Dudík, M., editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 627–635, Fort Lauderdale, FL, USA. PMLR.

Rozand, V., Lebon, F., Papaxanthis, C., and Lepers, R. (2015). Effect of mental fatigue on speed–accuracy trade-off. *Neuroscience*, 297:219 – 230.

Sacau, A., Laarni, J., and Hartmann, T. (2008). Influence of individual factors on presence. *Computers in Human Behavior*, 24(5):2255–2273.

Salcudean, S. E. and Yan, J. (1994). Towards a force-reflecting motion-scale system for microsurgery. In *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*, pages 2296–2301 vol.3.

Sasangohar, F., MacKenzie, I. S., and Scott, S. D. (2009). Evaluation of mouse and touch input for a tabletop display using fitts' reciprocal tapping task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 53(12):839–843.

Sathian, K. and Zangaladze, A. (2002). Feeling with the mind's eye: Contribution of visual cortex to tactile perception. *Behavioural brain research*, 135:127–32.

Savage-Knepshield, O. P., Hullinger, C.-C. D., Lund, P. R., Manning, C., Pierce, L., Seely, O., and Thomas, J. (2016). The challenges of measuring human performance in complex operational environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1):2053–2057.

Saxe, A., Nelli, S., and Summerfield, C. (2021). If deep learning is the answer, what is the question? *Nature Reviews Neuroscience*, 22(1):55–67.

Schaal, S. (1997). Learning from demonstration. In Mozer, M. C., Jordan, M., and Petsche, T., editors, *Advances in Neural Information Processing Systems*, volume 9. MIT Press.

Schäfer, L., Christianos, F., Hanna, J. P., and Albrecht, S. V. (2022). Decoupled reinforcement learning to stabilise intrinsically-motivated exploration. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '22, page 1146–1154, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

Schill, M. M., Gruber, F., and Buss, M. (2015). Quasi-direct nonprehensile catching with uncertain object states. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2468–2474.

Schoettler, G., Nair, A., Luo, J., Bahl, S., Ojea, J. A., Solowjow, E., and Levine, S. (2020). Deep reinforcement learning for industrial insertion tasks with visual inputs and natural rewards. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5548–5555. IEEE.

Schofield, W. N. (1976). Do children find movements which cross the body midline difficult? *Quarterly Journal of Experimental Psychology*, 28(4):571–582.

Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P. (2015). Trust region policy optimization. In Bach, F. and Blei, D., editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1889–1897, Lille, France. PMLR.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms.

Scribner, D. R. and Gombash, J. W. (1998). The effect of stereoscopic and wide field of view conditions on teleoperator performance. Technical report, Army Research Lab Aberdeen Proving Ground MD Human Research and Engineering.

Secoli, R., Milot, M.-H., Rosati, G., and Reinkensmeyer, D. J. (2010). Effect of visual distraction and auditory feedback on patient effort during robot-assisted movement training after stroke. In *Journal of NeuroEngineering and Rehabilitation*.

Shang, X., Kallmann, M., and Arif, A. S. (2019). Effects of correctness and suggestive feedback on learning with an autonomous virtual trainer. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*, IUI '19, pages 93–94, New York, NY, USA. ACM.

Shilling, R. D. and Shinn-Cunningham, B. (2002). Virtual auditory displays. In *Handbook of Virtual Environments*, pages 105–132. CRC Press.

Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review*, 20(1):21–53.

Simpson, B. D., Bolia, R. S., and Draper, M. H. (2013). Spatial audio display concepts supporting situation awareness for operators of unmanned aerial vehicles. *Human Performance, Situation Awareness, and Automation: Current Research and Trends HPSAA II, Volumes I and II*, 2:61.

Sinclair, M., Ofek, E., Gonzalez-Franco, M., and Holz, C. (2019). Capstancrunch: A haptic vr controller with user-supplied force feedback. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, page 815–829, New York, NY, USA. Association for Computing Machinery.

Slater, M., Linakis, V., Usoh, M., and Kooper, R. (1996). Immersion, presence and performance in virtual environments: An experiment with tri-dimensional chess. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '96, pages 163–172, New York, NY, USA. ACM.

Smyth, C. C. (2000). Indirect vision driving with fixed flat panel displays for near unity, wide, and extended fields of camera view. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(36):541–544.

So, R. H. and Griffin, M. J. (2000). Effects of a target movement direction cue on head-tracking performance. *Ergonomics*, 43(3):360–376.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., and Goodrich, M. (2006). Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, HRI '06, page 33–40, New York, NY, USA. Association for Computing Machinery.

Stoelen, M. F. and Akin, D. L. (2010). Assessment of fitts' law for quantifying combined rotational and translational movements. *Human Factors*, 52(1):63–77. PMID: 20653226.

Suzuki, Y. and Kobayashi, M. (2005). Air jet driven force feedback in virtual reality. *IEEE Computer Graphics and Applications*, 25(1):44–47.

Swan, J. E., Singh, G., and Ellis, S. R. (2015). Matching and reaching depth judgments with real and augmented reality targets. *IEEE Transactions on Visualization and Computer Graphics*, 21(11):1289–1298.

Tachi, S. (2010). *Telexistence*. G - Reference,Information and Interdisciplinary Subjects Series. World Scientific.

Tachi, S., Komoriya, K., Sawada, K., Nishiyama, T., Itoko, T., Kobayashi, M., and Inoue, K. (2003). Telexistence cockpit for humanoid robot control. *Advanced Robotics*, 17(3):199–217.

Taiga, A. A., Fedus, W., Machado, M. C., Courville, A., and Bellemare, M. G. (2020). On bonus based exploration methods in the arcade learning environment. In *International Conference on Learning Representations*.

Teather, R. J. and Stuerzlinger, W. (2007). Guidelines for 3d positioning techniques. In *Proceedings of the 2007 Conference on Future Play*, Future Play '07, page 61–68, New York, NY, USA. Association for Computing Machinery.

Teather, R. J. and Stuerzlinger, W. (2011). Pointing at 3d targets in a stereo head-tracked virtual environment. In *Proceedings of the 2011 IEEE Symposium on 3D User Interfaces*, 3DUI '11, page 87–94, USA. IEEE Computer Society.

Tee, K. P., Cheong, S., Li, J., and Ganesh, G. (2022). A framework for tool cognition in robots without prior tool learning or observation. *Nature Machine Intelligence*, 4(6):533–543.

Thor, M. and Manoonpong, P. (2022). Versatile modular neural locomotion control with fast learning. *Nature Machine Intelligence*, 4(2):169–179.

Thumser, Z. C., Slifkin, A. B., Beckler, D. T., and Marasco, P. D. (2018). Fitts' law in the control of isometric grip force with naturalistic targets. *Frontiers in Psychology*, 9:560.

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., and Abbeel, P. (2017). Domain

randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 23–30.

Torabi, F., Warnell, G., and Stone, P. (2018). Behavioral cloning from observation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, page 4950–4957. AAAI Press.

Triantafyllidis, E., Acero, F., Liu, Z., and Li, Z. (2023a). etriantafyllidis/roman: Roman v1.0.

Triantafyllidis, E., Acero, F., Liu, Z., and Li, Z. (2023b). Hybrid hierarchical learning for solving complex sequential tasks using the robotic manipulation network roman. *Nature Machine Intelligence*, 5(9):991–1005.

Triantafyllidis, E., Hu, W., McGreavy, C., and Li, Z. (2021). Metrics for 3d object pointing and manipulation in virtual reality: The introduction and validation of a novel approach in measuring human performance. *IEEE Robotics Automation Magazine*, pages 2–17.

Triantafyllidis, E. and Li, Z. (2021a). The challenges in modeling human performance in 3d space with fitts' law. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, New York, NY, USA. Association for Computing Machinery.

Triantafyllidis, E. and Li, Z. (2021b). Considerations and challenges of measuring operator performance in telepresence and teleoperation entailing mixed reality technologies. In *Evaluating User Experiences in Mixed Reality Workshop of the 2021 CHI Conference on Human Factors in Computing Systems*, Workshop CHI' 21, New York, NY, USA. Association for Computing Machinery.

Triantafyllidis, E., Mcgreavy, C., Gu, J., and Li, Z. (2020). Study of multimodal interfaces and the improvements on teleoperation. *IEEE Access*, 8:78213–78227.

Triantafyllidis, E., Yang, C., McGreavy, C., Hu, W., and Li, Z. (2020). *Robot intelligence for real-world applications*. IET Computing and Networks, UK.

Turk, M. (2014). Multimodal interaction: A review. *Pattern Recognition Letters*, 36:189 – 195.

Ullmer, B. and Ishii, H. (2000). Emerging frameworks for tangible user interfaces. *IBM Syst. J.*, 39(3–4):915–931.

Vaughan, J., Barany, D. A., Sali, A. W., Jax, S. A., and Rosenbaum, D. A. (2010). Extending fitts' law to three-dimensional obstacle-avoidance movements: support for the posture-based motion planning model. *Experimental brain research*, 207(1-2):133–138.

Vertegaal, R. (2008). A fitts law comparison of eye tracking and manual input in the selection of visual targets. In *Proceedings of the 10th International Conference on Multimodal Interfaces*, ICMI '08, page 241–248, New York, NY, USA. Association for Computing Machinery.

Vuskovic, V., Kauer, M., Szekely, G., and Reidy, M. (2000). Realistic force feedback for virtual reality based diagnostic surgery simulators. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*, volume 2, pages 1592–1598 vol.2.

Wang, Y. and MacKenzie, C. L. (1999). Object manipulation in virtual environments: Relative size matters. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, page 48–55, New York, NY, USA. Association for Computing Machinery.

Wang, Y., MacKenzie, C. L., Summers, V. A., and Booth, K. S. (1998). The structure of object transportation and orientation in human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, page 312–319, USA. ACM Press/Addison-Wesley Publishing Co.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. PhD thesis, King's College, Cambridge.

Weber, B., Sagardia, M., Hulin, T., and Preusche, C. (2013). Visual, vibrotactile, and force feedback of collisions in virtual environments: Effects on performance, mental workload and spatial orientation. In Shumaker, R., editor, *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments*, pages 241–250, Berlin, Heidelberg. Springer Berlin Heidelberg.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q.,

and Zhou, D. (2023). Chain-of-thought prompting elicits reasoning in large language models.

Welford, A. T. (1968). *Fundamentals of skill*. Methuen's manuals of modern psychology. Methuen, London.

Wellner, M., Schaufelberger, A., Zitzewitz, J. v., and Riener, R. (2008). Evaluation of visual and auditory feedback in virtual obstacle walking. *Presence: Teleoperators and Virtual Environments*, 17(5):512–524.

Witmer, B. G. and Kline, P. B. (1998). Judging perceived and traversed distance in virtual environments. *Presence: Teleoper. Virtual Environ.*, 7(2):144–167.

Wobbrock, J. O., Cutrell, E., Harada, S., and MacKenzie, I. S. (2008). An error model for pointing based on fitts' law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, page 1613–1622, New York, NY, USA. Association for Computing Machinery.

Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. (2011a). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 143–146, New York, NY, USA. ACM.

Wobbrock, J. O., Jansen, A., and Shinohara, K. (2011b). Modeling and predicting pointing errors in two dimensions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 1653–1656, New York, NY, USA. Association for Computing Machinery.

Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12(12):739–751.

Wu, Y., Fan, Y., Liang, P. P., Azaria, A., Li, Y., and Mitchell, T. M. (2023). Read and reap the rewards: Learning to play atari with the help of instruction manuals.

Yanco, H. A. and Drury, J. (2004). "where am i?" acquiring situation awareness using a remote robot platform. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, volume 3, pages 2835–2840 vol.3. IEEE.

Yang, C., Yuan, K., Zhu, Q., Yu, W., and Li, Z. (2020). Multi-expert learning of adaptive legged locomotion. *Science Robotics*, 5(49):eabb2174.

Yu, W., Gileadi, N., Fu, C., Kirmani, S., Lee, K.-H., Arenas, M. G., Chiang, H.-T. L., Erez, T., Hasenclever, L., Humplik, J., Ichter, B., Xiao, T., Xu, P., Zeng, A., Zhang, T., Heess, N., Sadigh, D., Tan, J., Tassa, Y., and Xia, F. (2023). Language to rewards for robotic skill synthesis.

Zaadnoordijk, L., Besold, T. R., and Cusack, R. (2022). Lessons from infant learning for unsupervised machine learning. *Nature Machine Intelligence*, 4(6):510–520.

Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications*, 10(1):3770.

Zaroff, C. M., Knutelska, M., and Frumkes, T. E. (2003). Variation in Stereoacuity: Normative Description, Fixation Disparity, and the Roles of Aging and Gender. *Investigative Ophthalmology & Visual Science*, 44(2):891–900.

Zeng, X., Hedge, A., and Guimbretiere, F. (2012). Fitts' law in 3d space with coordinated hand movements. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1):990–994.

Zhai, S., Milgram, P., and Buxton, W. (1996). The influence of muscle groups on performance of multiple degree-of-freedom input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, page 308–315, New York, NY, USA. Association for Computing Machinery.

Zhang, H., Starke, S., Komura, T., and Saito, J. (2018a). Mode-adaptive neural networks for quadruped motion control. *ACM Trans. Graph.*, 37(4).

Zhang, H., Ye, Y., Shiratori, T., and Komura, T. (2021). Manipnet: Neural manipulation synthesis with a hand-object spatial representation. *ACM Trans. Graph.*, 40(4).

Zhang, T., McCarthy, Z., Jow, O., Lee, D., Chen, X., Goldberg, K., and Abbeel, P. (2018b). Deep imitation learning for complex manipulation tasks from virtual reality teleoperation. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5628–5635.

Zheng, Y. (2013). An efficient algorithm for a grasp quality measure. *IEEE Transactions on Robotics*, 29(2):579–585.