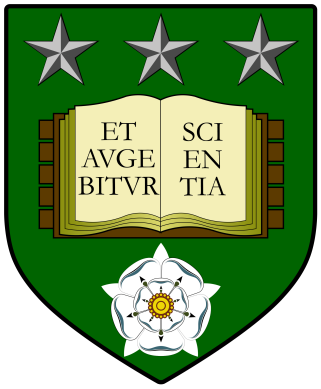
The Effect of Preoperative Simulation on Surgical Performance



Thomas William Pike

Submitted in accordance with the requirements for the degree of
Doctor of Medicine

The University of Leeds
School of Medicine
School of Psychology

March 2018

Declaration

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter 2 of this thesis has appeared in publication as follows: Pike TW, Pathak S, Mushtaq F, *et al.* A Systematic Examination of Preoperative Surgery Warm-up Routines. *Surg Endos.* 2017;31:2202. DOI: 10.1007/s00464-016-5218-x

The work in Chapters 4, 5, 6, 7, 8 and 9 of this thesis have appeared in publication as follows: Pike TW, Mushtaq F, Wilkie RM *et al.* How Should Surgeons Warm Up? An Experimental Psychology Approach. *JSS.* 2017;4:B6. DOI: 10.1102/2051-7726.2017.B006

The work in Chapter 10 of this thesis has appeared in publication as follows: Pike TW, Mushtaq F, Mann R *et al.* Operating List Composition and Surgical Performance. *Br J Surg.* 2108. DOI:10.1002/bjs.10804.

Acknowledgements

Firstly, I would like to thank my supervisors, Dr Faisal Mushtaq, Professor Richard Wilkie, Professor Mark Mon-Williams and Professor Peter Lodge, for their guidance, support and counsel, without which this thesis would not have been possible. I am especially grateful to Peter Lodge for his invaluable continuing mentorship and assistance in my surgical career, far above and beyond his role as as a supervisor.

I am indebted to Imogen Crook, Megan Wood, Flora Rees, Ella James-Brabham, Megan Skelton and Emily Robinson for their help and expertise in undertaking the experimental sections of this thesis.

I would also like to thank Zaed Hamady for starting me down the path that led to this thesis.

I am forever grateful for the continuing support of my family, especially my brother Oliver for his patient, indispensable coding advice.

Finally, I would like to thank my wife Lucy, my best friend and inspiration, without whom I would never accomplished anything.

For Charlotte

Abstract

Despite significant advances in both understanding and technology, complications of surgical care have become a major cause of death and disability worldwide. A substantial proportion of these complications are deemed preventable. It has been hypothesised that preparing surgeons to operate may improve performance and, by extension, patient outcomes. This thesis is concerned with three fundamental investigations into the effect of preparation (described as preoperative simulation) on surgical performance.

First is an investigation into the current understanding of preoperative simulation, which involves a systematic review of the literature. Broad support for preoperative simulation is demonstrated, however, the studies suffer from methodological shortcomings and lack theoretical grounding.

Building on this, a laparoscopic sequence learning task was developed to allow the investigation of preoperative simulation under controlled conditions. This was used in three controlled, randomised crossover trials. These experimental trials demonstrated that a simplified, relevant preoperative simulation can improve simulated laparoscopic performance. Exactly what form the preoperative simulation should take is determined by the nature of the task/operation being performed and is likely to reflect how that procedure was learnt. Additionally, preoperative simulation can alter an operators approach to completing a task, overriding a suboptimal but preferred method, to condition them to use a better method of completing the procedure.

Finally, the natural experiment of repeating a procedure during an operative list was used to explore the effect of preoperative simulation in clinical practice. An investigation of approximately a half million operations was conducted, which demonstrated that the order in which procedures are performed has a predictive relationship with operative duration (a surrogate for operative quality). This finding was relatively consistent across the thirty-five most common operations, and reinforced by the finding that switching procedures leads to significantly increased operative times.

These investigations support the view that preoperative simulation improves surgical performance.

Contents

List of Figures

List of Tables

Glossary

Roman Symbols

d	Cohen's d
$\bar{x}$	Mean
$\tilde{x}$	Median
R^2	Coefficient of Determination
s	Seconds
w_i	Akaike weight

Greek Symbols

χ^2	Chi-sqaured
ϵ	Greenhouse-Geisser Epsilon
η_p^2	Partial Eta-squared
μ	Population Mean
σ	Standard Deviation

Acronyms

AIC	Akaike's Information Criterion
AIC_c	Akaike's Information Criterion, adjusted for a small sample size
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
ASA-PS	American Society of Anaesthesiologists' Classification of Physical Status
$cf.$	*Conferre*
CI	Confidence Interval
CSS	Cascading Style Sheets
CSV	Comma-Separated Values
DPM	Dual Processor Model
EEG	Electroencephalograph
ER	Evidence Ratio
fMRI	Functional Magnetic Resonance Imaging
HTML	HyperText Markup Language
JS	JavaScript
LOS	Length of Stay

MIS	Minimally Invasive Surgery
ms	Milliseconds (10^{-3} seconds)
NHS	National Health service
NOTSS	Non-Technical Skills for Surgeons
NS	Not Significant
OGD	Oesophageal-gastro-duodenoscopy
OSATS	Objective Structure Assessment of Technical Skills
px	Pixels
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analysis
PSL	Primary Starting Location (of squares during the experimental trials)
PS	Preoperative Simulation
RCT	Randomised Control Trial
SDC	Sleep-dependent Consolidation
SRT	Serial Reaction Time
TSI	Task-set Inertia
WHO	World Health Organisation

Chapter 1

Introduction

'Complications of surgical care have become a major cause of death and disability worldwide.'

World Health Organisation, 2009

Approximately one major operation is performed annually for every twenty-five human beings alive; data from fifty-six countries in 2004 showed an estimated 187-281 million operations were performed worldwide [1]. This has significant implications for public health and well-being; it is almost double the number of childbirths per year [2] and is at least an order of magnitude more dangerous [3]. Complications following surgery are difficult to quantify due to the diversity of patients and procedures performed, but in industrialised countries, major complications following in-patient surgical procedures have been found to occur in 3-22% of all patients, with an associated mortality rate of 0.4-0.8% [4, 5]. Nearly half the adverse events that affected patients in these series were deemed to be preventable. In developing countries, studies suggest a mortality rate of 5-10% associated with major surgery [6–8].

Potentially avoidable complications are consequently responsible for a large proportion of preventable medical injuries and deaths globally [3]. In spite of significant improvements in surgical safety understanding, at least half of these events occur during surgical care [4, 5]. As deliberate, targeted public health interventions and educational programmes have led to dramatic improvements in maternal and neonatal survival, similar efforts may result in comparable improvements in surgical care and patient safety [9].

The World Health Organisation (WHO) has identified some of the major challenges to improving surgical safety [3]. One of these difficulties is the dearth of basic data.

Improvements in maternal and neonatal mortality have relied on the routine surveillance of obstetric care, which allows for the evaluation of interventions. Accepting that surgery presents quite a different challenge from childbirth, a similar robust system of monitoring does not exist globally in surgical care. Another reason identified by the WHO was the complexity of the entity being evaluated. For even the most simple operation to be successfully conducted, a prodigious number of component steps must be completed by an entire team of healthcare professionals and ancillary staff. However, simple interventions have been shown to have a significant impact on surgical safety and patient outcomes; the WHO Surgical Safety Checklist has been shown to approximately halve the total number of surgical complications and halve in-hospital mortality [10].

This work aims to explore some of the challenges faced when performing surgery, specifically related to the technical aspects of operating. Following the example of the WHO Surgical Safety Checklist, can simple preoperative interventions improve the practice of surgery? This introductory chapter describes some of the theoretical background to motor task learning and performance, on which all surgical performance is dependent. Surgery as a motoric skill is discussed in §1.1. Subsequently, in §1.2, motor skill learning and performance is discussed, with particular attention paid to sequence learning (§1.2.1) and adaptation (§1.2.2). The need for off-line processing and consolidation of learnt behaviour is reviewed in §1.3, with a summary provided in §1.4. A summary of the ground covered in this thesis is finally given in §1.5.

1.1 Surgery as a Motoric Skill

Surgery is a complicated, multifaceted undertaking, requiring the coordinated effort of multiple teams of healthcare professionals and supporting staff. Within the individual surgeon, a multitude of skills and attributes are required, which are encapsulated in *Good Surgical Practice* [11]. However, at its core, surgery relies upon practical, motoric skills.

In recent years, there has been increasing emphasis placed on the vital ancillary skills required by a surgeon, exemplified by the recognition of the importance of non-technical skills [12]. This recognition has lead to the development of assessment methods for Non-Technical Skills for Surgeons (NOTSS) [13], which is starting to be incorporated into postgraduate surgical training, as part of the major expansion in postgraduate assessment within the medical and surgical professions [14]. While recognising the importance of these non-technical skills, fundamentally, a surgeon must be able to

perform an operation. This ability to operate is contingent on the motor skills of the attending surgeon.

A motor skill can be broadly defined as an action which produces some measurable outcome of 'success' or 'quality' [15]. In contrast to habits or reflexes, motor skills require an intentional act [15]. The development of motor skills often requires extensive practice and the eventual evolution to skilled performance reflects the interplay of a variety of different factors. Coordination of the neuromuscular system, including elements such as biomechanics, postural control and reflexive constrains are required, as well as cognitive factors such as working memory, perception and characterisation [15]. This interaction of multiple factors can allow for the potentiation or impediment of the processes involved during both learning and performance of a task [16, 17]. For example, it has been demonstrated that both learning and performance of a task can be degraded by instructions that would intuitively seem to aid performance. Additionally, thinking too much about the task to be performed is particularly disadvantageous in stressful situations [18]. These findings support the accepted notion that educational or preparatory interventions may have a profound impact on skill performance. In the case of surgery, the archetypal high-stakes endeavour, any intervention should be soundly grounded in theory and have substantial experimental backing before being implemented in clinical practice. A core set of principles for motor skill learning and performance are discussed below.

1.2 Learning and Performance Behaviour

Early investigation of learning hypothesised lawful relationships between the properties of a task and its performance; Fitts' law [19] details the compromise between accuracy, speed and movement amplitude, while the Hick-Hyman law [20, 21] describes how information uncertainty increases reaction times. However, these laws have been shown to degrade with substantial practice [22], indicating motor performance is dependent on acquired knowledge about the task [15]. With practice, performance improvement is initially rapid, but the rate of improvement decreases as the practitioner becomes more skilled. Consequently, simple non-linear functions have long been used to describe the effects of practice across multiple tasks [23]. The supposed 'power law for practice' has become ubiquitous in describing this phenomena [24], however, as more fully explored later (see §3.3.1), the exponential function more accurately describes the effect of repeated practice on performance.

Regardless of how it is best modelled, initial rapid performance improvement fol-

lowed by subsequent slower improvement to eventual plateau is thought to be, at least partly, a consequence of the 'chunking' of information[1] [26]. Chunking combines the individual components of a task into functional units. On first encountering a new task, each movement is executed in isolation, i.e. a continuous movement between two pauses of the effector. As learning occurs, the individual elements become smoother and faster. There is also a progressive iteration where several contiguous elements are combined into 'chunks', resulting in a more efficient system with fewer and fewer chunks [27]. Classically, typing is used as an example of chunking, with the assumption that chunks are learnt hierarchically, i.e. chunks may be developed for representations of letter combinations, words or entire phrases [28].

Chunks have been considered representative of cognitive functions (spatial chunks), or as cooperations between forces, joints and muscles (motor chunks) [29–31]. The prominent Dual Processor Model (DPM) proposes a cognitive model of sequential motor skill development [29, 31]. The DPM suggests a cognitive and motor processor that execute discrete movement sequences. In early practice, the cognitive processor is responsible for translating each external stimulus into an associated response, and prompting the motor system to execute this response. In response to novel but explicitly known sequence, i.e. in the case of following instructions, a limited number of individual responses may be loaded into the motor buffer. The motor buffer is presumed to be part of working memory [32–34]. It is thought that, as short series of movements are repeatedly performed in succession, these sequences are integrated into a single representation; the motor chunk, Figure 1.1. Motor chunks eventually allow the cognitive processor to load the motor chunk(s) from long term memory into the motor processor as a single processing step, as if each motor chunk was a single response [32]. After loading chunks into the motor buffer, the cognitive processor signals the motor processor to begin reading the chunks and perform the movement series in a relatively autonomous manner. This process allows familiar sequences to be selected and executed in a rapid and precise fashion; a learnt skill. Consequently, as learning occurs, control of motor performance shifts from a general-purpose cognitive system to specialised motor system [27]. This allows the cognitive system to be 'freed up' to attend to other tasks while motor performance is executed by the motor system. In relation to laparoscopic surgery, this may allow experienced surgeons to dedicate more cognitive processing power to other aspects of operating, by requiring less dedicated cognition to be concerned with movement planning and motor control, when compared

[1] Though initially developed to explain the power law of practice, chunking is also applicable to an exponential model of learning, as it predicts a relative learning rate decrease to zero with practice [25].

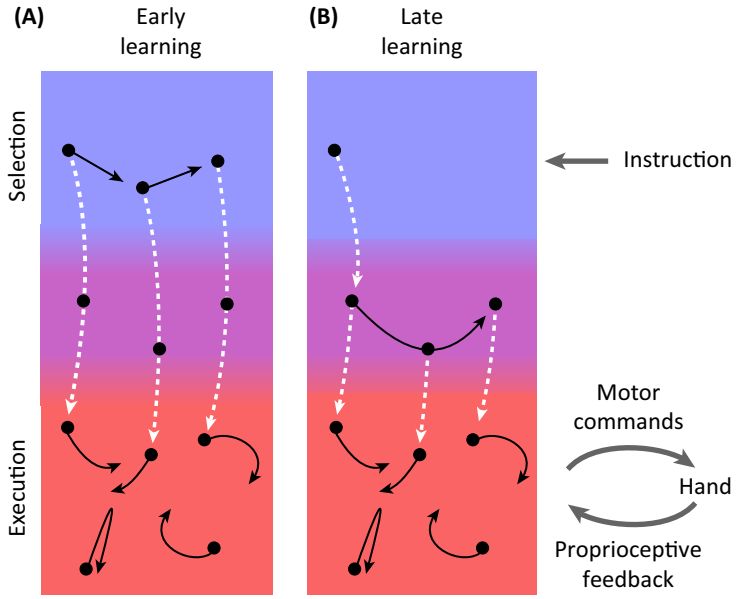


Fig. 1.1 Levels of Skill Learning, reproduced from Diedrichsen and Kornysheva [37]. In early learning (**A**), movement elements are activated (white broken line) from the cognitive processor (selection level, blue), which involves explicit processing of the task instruction. Skilled performance (**B**) seems to involve the formation of motor chunks at an intermediate level (purple), enabling easier recall and performance of complex movement combinations.

to junior surgeons. This may account for the improved intra-operative decision making senior surgeons demonstrate, despite potentially equivalent technical (motor) ability [35, 36].

Discrete chunks of motor performance are produced as a result of an efficiency computation trade-off [27]. Optimal motor performance requires the computation of muscle movements and joint angles to try and produce smooth trajectories and lower energetic requirements of execution (both of which are desirable in all movements). Importantly, these movements are calculated using dynamic programming and become exponentially harder to solve as the sequence of movements becomes longer [38–40]. Consequently, the computational cost of one extended sequence of movements is greater than the sum of shorter portions of the same sequence performed in series. Therefore, long sequences of movements are combined into a series of computationally simpler (and shorter) sequences; chunks. Experimentally, it has been shown that in the initial period of learning a motor skill, a large number of (short) chunks keeps the cost of computation low [27]. As learning occurs, movements within chunks are optimised, resulting in more efficient movements. Interestingly, Ramkumar *et al.* [27] suggested another method for the optimisation of movements over time; switching chunk orders to more efficient arrangement (for example, 3-3-2-2 versus 2-2-3-3), but found no evidence that

this occurs experimentally in primates. Although selecting chunk order for maximal efficiency across a set appears to offer an attractive strategy for developing optimal movements, it requires identifying the global minima; the smallest value across the entire domain of a function [41]. While optimisation can lead to a local minima, defining the global minima is dependent on finding the minima of each chunk individually, necessitating the generation of internal models of efficiency and re-learning chunk order from one trial to the next, which is likely to introduce additional demands on working memory. Additionally, the most efficient order for a set of chunks will be dependent on the structure of the constituent chunks, meaning that a change to the order of an individual chunk is likely to necessitate a change to overall order. Consequently, optimisation within chunks, even if better potentially unexplored chunk orders exist, appears to be an effectively simple solution.

1.2.1 Sequence Learning

Sequences of information and/or actions are encountered in almost all tasks; from sequencing the compound sounds of speech, to the individual movements when typing, to the sequence of actions required to perform minimally-invasive surgery. Serial reaction time (SRT) tasks have long been used to study sequence learning. This commonly requires a participant to push a key in correspondence to a visual signal, similar to a simplified form of typing. Reaction time has been shown to reduce when responding to predictable stimuli, and provides a measure of sequence learning [42]. The flexibility of a learnt skill can be examined by assessing performance of the skill in a novel environment, using a transfer study design. In transfer tasks, the performance of a task with the finger movements of one hand have been shown to effectively transfer to arm movements, finger movements with the opposite hand and even verbal responses [43, 44], but interestingly, the degree of transfer is dependent on the amount of practice [15]. For example, inter-manual transfer has been shown to occur after one hour of practice, but not after five weeks of training [45], suggesting extended practice ties a skill to a particular mode of performance. This has significant potential implications for the development of minimally-invasive surgery training regimes. With the exponential increase in use of simulation in surgical education [46], trainees have a much greater exposure to predictable stimuli (simulators replicate exactly the same operations again and again but are unable to replicate the diversity found in the population), when compared to previous generations of surgical trainees. While simulator training has been shown to be beneficial [47, 48], it may be possible to overtrain on a simulator, resulting in a lack of ability to transfer to real-world operating. This may also explain why

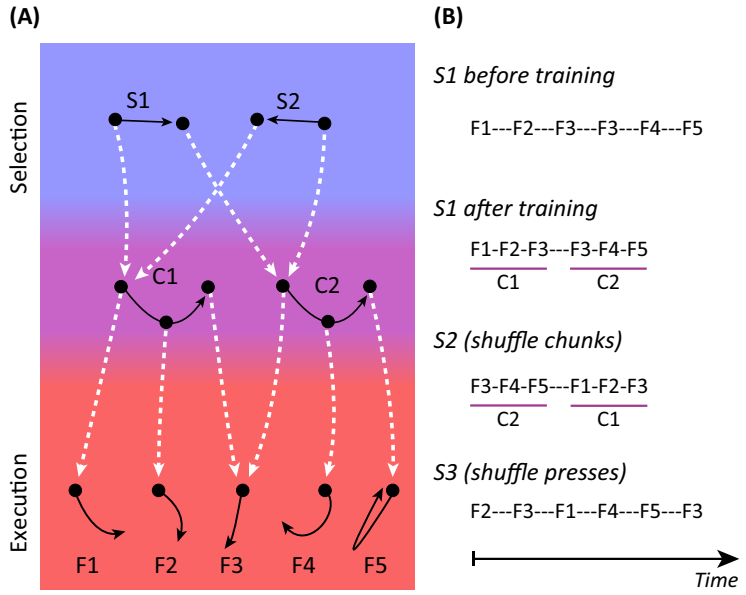


Fig. 1.2 Movement Chunking, reproduced from Diedrichsen and Kornysheva [37]. (**A**) Triggering individual chunks (C1, C2) from the selection/cognitive level results in the performance of individual movement elements (F1-F5). (**B**) Training using the S1 sequence leads to faster and move accurate in novel sequences when acquired chunks are preserved (S2), but not when the chunks are broken up (S3).

proficiency in one minimally-invasive technique does not necessarily correlate with proficiency in another; operative outcomes are dependent on a surgeon's experience with a single procedure [49–51]. This hypothesis is supported by the finding that participants demonstrated different functional magnetic resonance imaging (fMRI) activity after one hour and five weeks of training when repeatedly performing a motor task [45], demonstrating a change within the motor cortex that suggests extended practice may anchor a skill to a particular form of execution. However, such findings do not take into account the performance of an expert surgeon who is able to accommodate variations from the norm and operate in novel situations (see Lodge *et al.* [52] as an example). Consequently, consolidation may initially limit flexibility, producing performance with minimal variation and stability, but eventually allow a plasticity of 'chunks' (discussed above) that can be flexibly varied and combined to allow performance under novel conditions [37, 53]; Figure 1.2. The reconfiguration of learnt chunks may be employed by experts to perform at high levels in novel environments. Such performance would necessitate the recruitment of additional cognitive resources, but, as discussed above, expert performance is less cognitively demanding than novice performance, permitting the use of additional cognitive processing for other tasks such as chunk reconfiguration.

Variation in ability to transfer skills is likely to reflect the involvement of differ-

ent learning pathways with differing timescales. Early during the learning process, 'fast' learning processes may produce abstract representations of the task that can easily generalise across different effectors and domains. As learning continues, 'slow' effector-specific processes become predominant. Although more able to generalise, faster systems may be disrupted by additional, unrelated task requirements, whereas slower systems are likely to be more automated and consequently better at mitigating the effects of distraction [15]. This is demonstrated in practice by the trainee who is able to perform a certain task on a box-trainer, but struggles to do so when operating in theatre, compared to the expert surgeon who is able to perform a task under both controlled conditions (simulation) and during times of significant distraction (operating in theatre).

1.2.2 Adaptation

All minimally invasive surgery forces the surgeon to adapt to a visuospatial transformation. During laparoscopic abdominal surgery, the surgeon views a two-dimensional representation of the three-dimensional abdominal cavity which is positioned away from the field of operating and all movements are perturbed, sometimes reversed, by the fulcrum effect of laparoscopic ports. Once again, as above, parallel systems of learning, operating on differing timescales, have been proposed to account for the experimental findings of adaptation tasks. During visuomotor adaptation, normal visual input / motor output mappings are modified, for example, a participant's visual feedback may be rotated around a fixed point or the participant may experience perturbing forces during performance. Adaptations to such manipulations typically demonstrate exponential decay over repeated trials [23], mirroring the learning trials discussed above. This phenomenon can be explained as the result of multiple systems of learning working in parallel [54]. 'Fast' learning systems result in rapid learning, but also rapid decay. 'Slow' systems require a greater number of trials to achieve the same degree of adaptation, but are relatively stable and resistant to degradation, which again mirrors the sequence learning trials discussed above.

Awareness of the need to adapt to a visuomotor transformation has produced interesting observed effects. Participants have been shown to be able to adapt to a forty-five degree visuomotor rotation immediately, after being told to aim for a strategic target placed forty-five degrees in the opposite direction to the rotation [55]. Paradoxically, over time the participants' accuracy declined as greater rotations than required were produced. It was suggested that this apparent contradiction is again due to differing learning systems operating on different information. Operating on strategic instruc-

tions (i.e. "aim for *this* target"), one system may generate a motor plan that involves aiming to a strategic target forty-five degrees from the actual target. Concurrently, another system may try and calibrate movement based on the difference between the predicted and perceived endpoints, while lacking information regarding the actual target. Combining these two systems will result in a compound movement away from the actual target towards the strategic target. This has led to the hypothesis that, when trying to account for visuomotor transformations, a participant can engage control systems under explicit instruction, which may assist in rapid learning, but hamper expert performance of the task [17]. Such findings have an implication for surgical education and preoperative preparatory regimes - explicit instructions may aid the development of surgical skills, but impair expert performance.

1.3 Off-line Processing and Consolidation

In the United Kingdom, surgical trainees' are unlikely to finish higher surgical training and reach consultant (independent practitioner) level before the age of thirty-five [56], though this does not mean that learning ceases at this point. In order to continue to learn over such protracted periods of time, a balance between flexibility and longevity is required. This is again provided by multiple learning systems operating over different timescales. 'Fast' learning mechanisms are able to generate new representations when confronted with novel tasks or environments, but require greater attention and are comparatively cognitively demanding. 'Slow' learning systems are able to decrease cognitive load and retain learnt skills despite experiencing novel learning [15]. Skills retained over prolonged periods of time (a hallmark of motor skills in general) are considered 'consolidated'. There is increasing evidence that, for at least some aspects, consolidation has to occur 'off-line', when the task is not actively being performed [57–60]. For some forms of learning, this consolidation is dependent on sleep to occur (particularly 'fast' systems), while others will occur after the passage of time.

Consolidation of learning can be difficult to examine, but can be demonstrated as savings in relearning; when a task is repeated after an extended break, performance improves faster then during the initial acquisition of the skill [15]. Additionally, consolidation can be evident in resistance to interference during task performance. After initial training, resistance is more pronounced the following day (i.e. after sleep), whereas after prolonged training similar resistance to interference is demonstrated after a five-minute break or the following day [61]. It is possible that allowing a period of 'washout' after initial trials may be useful in consolidating more abstract repre-

Table 1.1 Characteristics of Multiple Systems for Motor Learning, reproduced from Clark and Ivry [15].

'Fast' Learning Systems (I)	'Slow' Learning Systems (II)
Large amount of learning per trial that saturates quickly (high gain)	Small, incremental amount of learning per trial (low gain)
Requires extra time, cognitive resources for processing	Learns automatically without effort
Required for contextual learning	Unimodal or modular learning
Accessible to awareness and conscious intention	Impenetrable to awareness, operates independent of conscious strategies
Consolidation processes are enhanced during sleep	Consolidates off-line with the simple passage of time
Ready transfer to related tasks	Effector-specific and inflexible

sentations of the task, generated from 'fast' learning systems, while further training shifts performance into more specific (less plastic) representations, derived from slower systems, which are more consolidated and less affected by interference.

Practically, the development of surgical skills occurs over many years, allowing ample opportunity for off-line processing and consolidation. However, the need for the off-line processing and consolidation of skills needs to be considered in experimental designs examining the effects of surgical education and the potentiation of learnt task performance.

1.4 Summary of Motor Skill Learning Theory and Relation to Laparoscopic Surgery

The theoretical evidence presented above suggests that multiple independent systems, operating over different time scales, support motor learning; see Table 1.1. Whilst it has been suggested that more precise characterisation of motor learning mechanisms will be provided in neural terms [15], the above delineative provides a useful framework which can be utilised in developing methods to aid both learning and learnt performance.

Minimally-invasive surgery can be thought of as a (albeit complicated, multidimensional) sequence learning task, as a surgeon must perform a series of tasks, in a predetermined order, to successfully complete an operation. During minimally-invasive surgery (*cf.* open surgery), all of the task must be completed while coping with a visuospatial transformation of normal visual input and motor output, which places greater demand on a surgeon's cognitive functions.

There has been a proliferation of minimally-invasive surgery in recent years and, at the same time, a reduction in training time for surgeons [62–64]. As surgery remains a

major cause of death and disability worldwide, and half of all the adverse events that affect patients occur during the practice of surgery, could an intervention that improves a surgeon's operative performance result in better clinical outcomes? The WHO Surgical Safety Checklist [10] has demonstrated that simple preoperative interventions can have a significant impact on patient outcomes. Might surgeons benefit from some form of 'warm-up', which is utilised extensively in other high-stakes motoric performances, such as elite sport? In such domains, warming up, both physically and 'mentally', has been shown to benefit performance [65, 66]. This concept has recently begun to be investigated in surgery [67, 68], though such investigations are nascent, as discussed in Chapter 2. As such warm-ups in surgery most commonly employ a surgical simulator, the practice of warming up prior to operating is hereafter referred to as preoperative simulation.

This work aims to explore the effect of preoperative simulation on surgical performance. Three approaches are utilised to try and achieve this goal; firstly, a systematic evaluation of the current understanding is undertaken. Secondly, experimentation under strictly controlled conditions is performed. In order to achieve rigorous control over the varied extraneous factors than can affect surgical performance, the experiments performed in this thesis are necessarily removed from the reality of undertaking surgery. As discussed in Chapter 2, this is not always the case when surgical technologies are being developed or evaluated. However, without stringent controls, at least during the initial development and assessment of a supposedly useful intervention, the precise beneficial components of any technology are likely to be lost in noise. Once a principle has been demonstrated under controlled conditions, evaluation under less well controlled, more true to life scenarios can occur. Such an approach is analogous to the initial basic science behind any novel pharmaceutical development. Finally, in order to glean some insight into the effect of preoperative simulation in clinical practice, the natural experiment of repeating a procedure on a operating list is explored.

1.5 Structure of Thesis

Chapter 2 summarises the current literature that examines the effect of preoperative simulation on surgical performance. A systematic review was conduced according to PRISMA guidelines [69], which identified thirteen relevant articles. The results of these studies, including a frank appraisal of methodological shortcomings, are discussed.

Chapter 3 is the first of seven chapters of original work. To try and identify the

effective component(s) of a beneficial preoperative simulation routine, a bespoke laparoscopic sequence learning task was developed (§3.1), which can produce detailed metrics of performance (§3.1.1). The overarching structure of subsequent experimental work utilising this program is subsequently discussed (§3.3).

Chapters 4 & 5 detail the first of three experiments utilising the laparoscopic sequence learning task (Experimentα). Chapter 4 reports the specific task developed for this experiment (§4.1), as well as the methods (§4.2) and results (§4.4) of learning trials, designed to teach participants the experimental task. A discussion of these result is subsequently presented (§4.5). Following, Chapter 5 presents the methods (§5.1), results (§5.2) and a discussion (§5.3) of the effect of preoperative simulation on performance of the Experimentα task.

Chapters 6 & 7 provides the results from Experimentβ, which follows a very similar experimental design to Experimentα, but examines the effect of learning (Chapter 6) and preoperative simulation (Chapters 7) on a different task, developed using the laparoscopic sequence learning task program presented in Chapter 3. A synthesis of the results of both Experiment α and β is discussed in §7.4.

Chapters 8 & 9 are the final experimental chapters and report the results of Experimentγ, which again following a similar experimental structure to the previous two experiments. A consolidation of all experimental results, including the shared characteristics of effective preoperative simulations, is presented in §9.3. The potential implications for the use of preoperative simulation in clinical practice are developed in §9.4.

Chapter 10 outlines a big data investigation into the natural experiment of repeating a procedure on a single operating list, allowing an exploration of the effect of preoperative simulation in current clinical practice. Following on from preliminary analysis utilising local data (§10.1), the analysis of 478,713 operations from thirty-eight Spire Healthcare hospitals across the UK is presented (§10.2). These results demonstrate, for the first time, that the order in which procedures are performed has a predictive relationship with operation duration (§10.3). A discussion of this finding, including its application to preoperative simulation is discussed in §10.4.

Finally, **Chapter 11** summarises the main results of this thesis and discusses the various paths that have opened for further work.

Chapter 2

Systematic Review of Current Evidence

To understand the processes driving an effective preoperative simulation, a systematic review of the literature on preparation and surgical success was conducted. After identifying studies examining the efficacy of preoperative simulation, the task characteristics of a successful preoperative simulation and the performance metrics that are modulated by this process were considered.

2.1 Methods

A search strategy according to the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidance [69] was developed. An electronic search of relevant databases (Cochrane Library (1995-), PubMed, PsycINFO (1967-), ERIC (1964-) and Google Scholar) was conducted utilising the following key words: "Surgery", "Laparoscop*", "Minimally Invasive", "Simulat*", "Educat*", "Technolog*", "Warm-up", Warm up", "Preparation", "Planning", "Rehearsal", "Mental Rehearsal", "Cognitive", "Decision Making", "Decision", "Outcome", "Performance", "Preoperative" and "Preoperative". Key words were grouped using "AND" or "OR" terms. Bibliographies of relevant studies and the "related articles" link in PubMed were used to identify any additional studies. All citations and abstracts identified were thoroughly reviewed. The last date for this search was 1st May 2015.

2.1.1 Inclusion Criteria

All included studies analysed the effect that a preoperative simulation had on subsequent surgical performance (simulated or real-life). Studies were restricted to those

that examined a deliberate intervention prior to an operation or procedure as opposed to a training regime or educational programme. There was no restriction applied to the type of skills trained or assessed. All study designs were considered for inclusion.

2.1.2 Exclusion Criteria

All citations published only as an abstract or unpublished report were excluded from further analysis. All studies were carefully evaluated for duplication or overlapping data and such reports removed.

2.1.3 Outcome Measures

The primary outcome of interest was surgical performance, however defined. Of secondary interest were the outcome measures reported by each study.

2.1.4 Study Selection

The search was performed according to the strategy described above. The identified abstracts were reviewed those that did not meet the inclusion criteria and excluded. If no abstract was available or the abstract did not contain adequate information, the full article was reviewed.

2.1.5 Data Extraction

Data extraction was performed using a standardised pro forma. The following parameters were recorded: study characteristics (first author, year of publication, place of publication), population characteristics and outcomes of interest.

2.1.6 Risk of Bias Assessment

The method for objectively assessing the risk of bias of included studies depended on the type of study. Randomised control trials were reviewed using the Cochrane risk of bias tool [70] while cross-over trials were analysed using a modified version of a tool developed by Mills *et al.* [71], specifically for reviewing cross-over trials.

2.1.7 Statistical Analysis

The heterogeneity of included studies prevented a quantitative synthesis of reported outcomes.

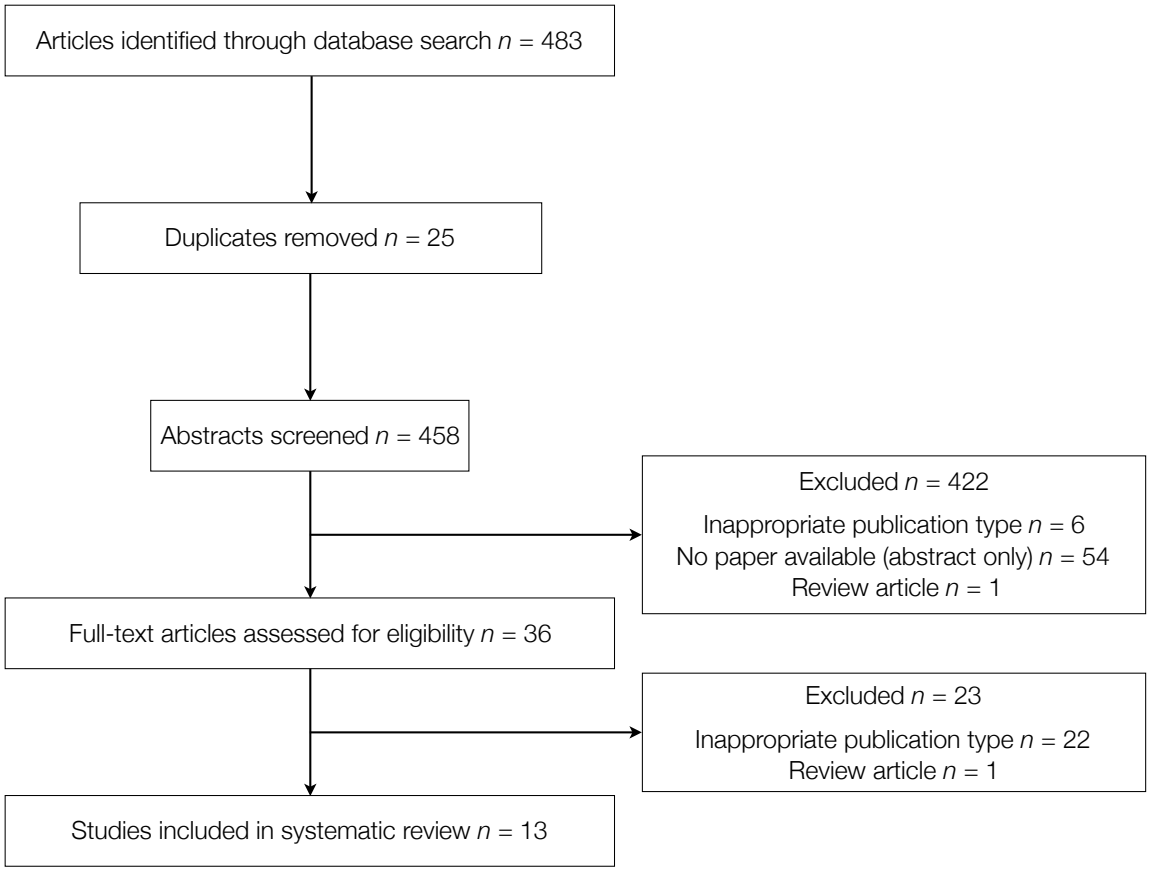


Fig. 2.1 PRISMA flowchart depicting the search strategy and selection of articles for the review.

2.2 Results

Four hundred and eighty three articles were identified by the search strategy described above. Following a review of abstracts, full articles and references, 13 studies were included in this systematic review: Figure 2.1.

2.2.1 Study Characteristics

Four randomised control trials (RCTs) [72–75] and four randomised cross-over studies [76–79] were included, all of which reviewed operative outcomes following a practice of technical skills, prior to an operation versus no practice. A further four studies were case studies, two of which compared a technical skills practice to no practice [80, 81] and two of which [82, 83] did not contain a control group. One RCT [84] examined the effect of mental practice prior to an operation on subsequent laparoscopic performance.

Eight of the studies [72, 74, 75, 79–81, 83, 84] examined the effect of preoperative

simulation on general surgery procedures, three looked at obstetrics and gynaecological procedures [73, 76, 82] and the last two examined the effect of preoperative simulation on endovascular [77] and urological [78] procedures. Four of the studies [72, 78–80] examined outcomes in real patients, the other nine [73–77, 81–84] reviewed simulated outcomes.

Three mediums of simulation were employed by the included studies; in seven studies a virtual-reality simulator [75, 77–80, 83], in four a laparoscopic box trainer [72, 73, 76, 82], and in two video-games [74, 81] were used as preoperative simulation. Though various forms of simulation were employed, there was general concordance across the studies as to what constituted preoperative simulation. All studies, except one [84], used a similar or simplified motor task as preoperative simulation before performing the assessed task.

2.2.2 Assessment of Bias

There was significant variability in the quality of studies included. Only one study [75] was judged to be at low risk of bias: Appendix A. Five studies [72, 73, 75, 76, 79] were found to be at low risk of randomisation bias, with explicit detailing of the methods of randomisation and allocation concealment employed. Two studies [72, 75] reported *a priori* power calculation, but one of these studies [72] calculated that a significantly larger number of participants would be required than were actually recruited. The sample sizes within each study were generally modest, with only one RCT or cross-over study [75] reporting more than 20 participants per group. The included case studies could not be objectively assessed by the methods used to review the RCT and cross-over studies, but each demonstrated methodological shortcomings, as discussed below.

2.2.3 Reported Outcomes

The included articles report 103 different outcome metrics, often combined to form a compound score. A summary of the main findings of each study is detailed in Appendix B. Twelve [73–84] of the thirteen manuscripts concluded that a preoperative simulation improves subsequent surgical performance. No study found a preoperative simulation to have a detrimental effect on surgical performance or suggested any negative aspect of preoperative simulation.

2.2.4 Studies Reporting Global Rating Score

The most often reported outcome was the effect of preoperative simulation on a global rating score of performance. Seven of the included studies [72, 73, 77–80, 84] reported this outcome metric, defined as a summary of objective assessment parameters by an expert examiner. Nine global rating scales [72, 73, 75, 85–91] were employed in the seven studies, with all but two studies reporting a different global rating scale. In a majority of studies validated global rating assessments were used. In three studies [72, 78, 80] a modification of a previously published global rating scale was employed.

In two RCTs [73, 84] and one cross-over trial [79], the authors reported unequivocally that preoperative simulation improves subsequent real-world operative performance, as assessed by a global rating scale. Two cross-over trials [77, 79] report ambiguous findings for the effect of preoperative simulation on surgical performance; one study [77] reported a significant effect as measured by one global rating scale, but no effect according to another also-reported scale. The other study [78] found a significant improvement in one assessed task, but not another. One RCT [72] and one cross-over trial [80] found that preoperative simulation had no effect on subsequent performance, as judged by a global rating scale.

2.2.5 Studies Reporting Performance Time

Five of the included studies [74, 75, 77, 81, 83] reported 'pure' performance time, defined as the time taken to perform an assessed task. Those articles that reported duration as part of a global rating scale were not included as such studies have been discussed above. The authors in one RCT [75] and one case study [76] reported that preoperative simulation shortens subsequent performance time in a simulated environment. In one cross-over trial [77] and one case study [83] equivocal results were reported, with preoperative simulation reducing the time of some performance metrics, but not all. Finally, in only one RCT [74] did preoperative simulation not affect the time taken to perform simulated laparoscopic surgery.

2.2.6 Studies Reporting Time-based Score

In three studies [76, 81, 82] a time-based score was reported, either in combination with errors made (resulting in a time penalty) or as the number of occasions a task was performed within a set time. The authors in one case study [82] found that preoperative simulation increases the number of times a laparoscopic task can be performed within a set time period. The authors in another case study [81] found that preoperative

simulation reduces time taken and errors made during the placement intracorporeal sutures, but not the time taken and errors made during two other laparoscopic tasks. Finally, in one cross-over study [76] it was found that preoperative simulation did not improve simulated laparoscopic performance as assessed by a time-based score.

2.2.7 Studies Reporting Cognitive Performance

In four studies [74, 75, 78, 83] the effect of preoperative simulation on participants' cognitive performance, most commonly defined as errors made during a procedure (determined by a simulator or expert assessor) but also based on electroencephalograph (EEG) readings during performance, was reported. The authors of one RCT [74] and one case study [83] found that preoperative simulation significantly reduced the number of errors that occurred during simulated laparoscopic performance. Conversely, another RCT [75] found that preoperative simulation did not affect the number of cognitive errors made. In one cross-over study [78] it was found that preoperative simulation improved attention, reduced distraction / drowsiness and reduced mental workload when compared to no simulation.

2.2.8 Studies Reporting Simulator-generated Metrics

The authors of four studies [74, 75, 78, 83] reported outcome metrics generated by the laparoscopic simulator used during their experiments. Hand and tool movement smoothness and instrument path length were reported, but there was no concordance across the studies; some reported significant results in certain outcome metrics while others did not.

2.2.9 Studies Reporting Mental Imagery

One RCT [84] reported participants' mental imagery following mental practice (experimental group) or an online academic activity (control group) immediately prior to performing a simulated laparoscopic cholecystectomy. Mental practice was defined as "the cognitive rehearsal of a task in the absence of overt physical movement" [84]. The authors reported undertaking structured mental practice significantly improved participants' mental imagery of a procedure.

2.2.10 Studies Reporting Participants' Perception

The authors of one cross-over study [77] explored participants' perception of how useful they found the preoperative simulation and whether participants felt preoperative simulation improved their subsequent performance. This was assessed by a questionnaire utilising a five-point Likert scale. Participants reported that they felt patient-specific simulation to be more helpful than generic simulation, which was more useful than no simulation. Participants also reported that they felt that patient-specific simulation helped with decision making, improved safety, increased their confidence levels and resulted in reduced preoperative anxiety (of the operator).

2.2.11 Studies Examining Outcomes in Real Patients

In four studies [72, 78–80], the effect of preoperative simulation on real patients was examined. Three of these studies [78–80] concluded that pre-operative simulation improves real operative outcomes. These studies assessed participants' performance during laparoscopic renal surgery (mobilisation of the colon and intracorporeal suturing and knot tying) [78] and laparoscopic cholecystectomy [79, 80], reporting improvements in cognitive and psychomotor performance [78], and some global rating score metrics [78–80]. Weston *et al.* [72] found preoperative simulation to have no effect on subsequent performance. However, as discussed above, Weston *et al.* performed an a priori power calculation that demonstrated a larger number of participants than were actually recruited would be required to achieve statistical significance. Consequently, the absence of a significant result may be due to a lack of statistical power in the study.

2.2.12 Underlying Processes Examined by the Included Studies

In order to explore the underlying mechanisms for the observed performance improvements through preoperative simulation, a rudimentary analysis of the tasks performed was conducted, using the frameworks discussed in Chapter 1; see Appendix C. Whilst acknowledging that these systems work in concert, and although necessarily speculative in nature, investigation into the degree of engagement of each system could assist with the future development of optimal preoperative simulation interventions. Thus, we categorised the preoperative simulation routines employed in each study into those more likely to engage 'fast' learning processes (e.g. motor practice) and those more likely to utilise 'slow' learning processes (e.g. those deliberated, overtly effortful and requiring cognitive control). The majority of included studies [72–83] indicate a greater

degree of 'fast' processes in their preoperative simulation routines. In most of these cases, a simplified simulated task (*cf.* the assessment task) was used to prepare the participants for surgery (real-life or simulated). Two studies [77, 80] are likely to have engaged both 'fast' and 'slow' learning processes and one study [84] relied more heavily on 'slow' learning processes.

2.3 Discussion

This systematic review was conducted to investigate the utility of preoperative simulation as a means of improving decision-making and performance in surgery. All but one of the studies included in this review concluded that preoperative simulation significantly improves subsequent surgical performance. However, the results presented above, when synthesised, do not appear to present such a coherent picture. One of the reasons for this discrepancy is the number of outcome metrics used in each study; see Appendix B. Only four studies [72, 73, 79, 82] report unequivocal results i.e. there is concordance between all reported outcome measures within the studies. Three of these studies [73, 79, 82] concluded that preoperative simulation improves surgical performance. One study [72] reported that preoperative simulation does not affect subsequent performance. However, as noted earlier, this study recruited a significantly smaller sample size than the authors calculated would be required. The nine other studies [74–78, 80, 81, 83] included in this review reported significant results in some, but not all, recorded outcome measures. All of these studies concluded that preoperative simulation improves surgical performance, but only two studies [75, 78] include an explanation as to why significant results are prioritised over non-significant results. This selective reporting of significant outcome measures may bias the conclusions drawn from these studies. This is an issue that generalises; a consensus opinion on outcome reporting is imperative to allow effective meta-analysis of results and allow a high quality evidence base to be developed. Surgical education and training needs to follow the example being set by clinical research by agreeing a set of standardised outcomes to report [90, 92–94]. Whilst this issue is one that is beyond the scope of the current review, it appears that consensus is particularly necessary in the assessment of simulated surgical skills. Surgery performed on real patients can be assessed by reviewing patient outcomes (although none of the included studies reported such outcomes), which must be considered the gold-standard of outcome reporting. However, simulation-based research often relies on outcomes of convenience. For example, one of the most frequently reported outcome metrics in this systematic review was performance time. Whilst it

is recognised that expert performance is faster than novice performance [47, 48], the converse is not necessarily true, i.e. faster performance does not necessarily confer better quality surgery [95]. The same applies to simulator-generated metrics; for example, while experts tend to have smoother movements, having smoother movements does not necessarily mean the operator is an expert. Consequently, such metrics, particularly when reported without additional objective or subjective data, can only be interpreted with considerable caution. Despite this, the majority of reported outcomes demonstrate that preoperative simulation does have a beneficial effect on subsequent surgical performance, in both simulated and real-patient environments

In this review, studies were included irrespective of the type of surgical skill being examined. The heterogeneity of the included studies can be viewed as a strength of this review as the generic concept of preoperative simulation can be explored across multiple surgical specialties, using a variety of assessment methods. Conversely, the disparity between studies and the number of different outcome metrics used, prevent a quantitative synthesis of reported outcomes. In addition, because of the paucity of studies in the literature, designs that are often excluded from systematic review have been included.

The majority of included studies examined the effect on surgical performance of a simplified motoric task. However, the included studies were conducted without any reference to a theoretical framework, as evidence by the narrow focus of preparatory procedures employed. As outlined in Chapter 1, motor learning and performance can be understood as an interaction between 'fast' and 'slow' systems. The studies reviewed here predominantly focused on more automated behaviours ('fast') at the neglect of controlled cognitive processes ('slow'). We speculate that interventions relying on both are likely to produce greater benefit than focusing on a single process alone [96].

In conclusion, the majority of reported outcomes demonstrate that preoperative simulation does have a significantly beneficial effect on subsequent surgical performance, in both simulated and real-patient environments. Importantly, no study reported preoperative simulation had a detrimental effect on subsequent performance. Thus, it appears that surgeons may benefit from engaging in formalised preparation routines before carrying out an operation.

Chapter 3

ESOX: A Programmable Laparoscopic Sequence Learning Task

The evidence presented in Chapter 2 suggests that preoperative simulation has a significant impact on operative performance. However, all of the studies included in the systematic review demonstrated methodological shortcomings which may confound their conclusions. Thus, an exploration of the fundamental underpinnings of preoperative simulation was undertaken, through experimentation under controlled conditions, to try to establish the components underlying a beneficial preoperative routine. In this chapter, the development of a bespoke computer program - ESOX - is described.

3.1 The ESOX program

A computer program, ESOX, was developed by the author to allow exploration of factors that affect the performance of laparoscopic surgery, under strictly controlled conditions. ESOX was created using HyperText Markup Language (HTML) 5, and utilising additional Cascading Style Sheets (CSS) and JavaScript (JS) code. Consequently, ESOX can run on any device that supports a web browser.

ESOX was designed to allow the development and rapid deployment of sequence-learning tasks. It generates a user-specified x by y grid which can then be populated by numbered, coloured squares, as determined by an accompanying comma-separated value file (.csv); see Figure 3.1 as an example. Participants are asked to move to each square, in order. Once the correct square is reached it 'disappears'. Participants are given feedback about their performance in the form of time taken (seconds) and

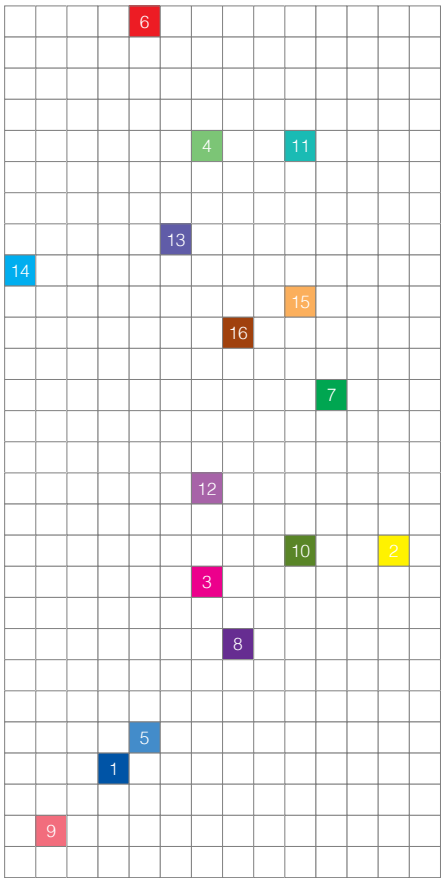


Fig. 3.1 Example of ESOX output (Experimentα)

distance travelled (pixels) at the end of the task. ESOX is also able to vary the position of squares within the grid between trials randomly by automatically applying a Fisher-Yates shuffle [97] every time the program is loaded[1], if required. In addition, it is able to use 'fuzzy' spatial location parameters, altering the position of squares by a predetermined amount (i.e. ± 2 squares in the x or y direction).

ESOX records the position of the cursor every 10 milliseconds (ms) the program is running, as well as when each individual coloured square is reached. This information is automatically downloaded as a CSV file every time the task is completed. Post-performance processing of these files allows for detailed metrics of performance to be calculated.

In order to simulate laparoscopic surgery, ESOX is designed to run on a touch-enabled tablet computer contained within a laparoscopic box trainer. The program is then controlled using a laparoscopic stylus, with information being displayed on a wide-screen monitor in front of participants (participants are unable to see within the box trainer): Figure 3.2.

[1] A Fisher-Yates shuffle elegantly generates an unbiased random permutation of a finite set.

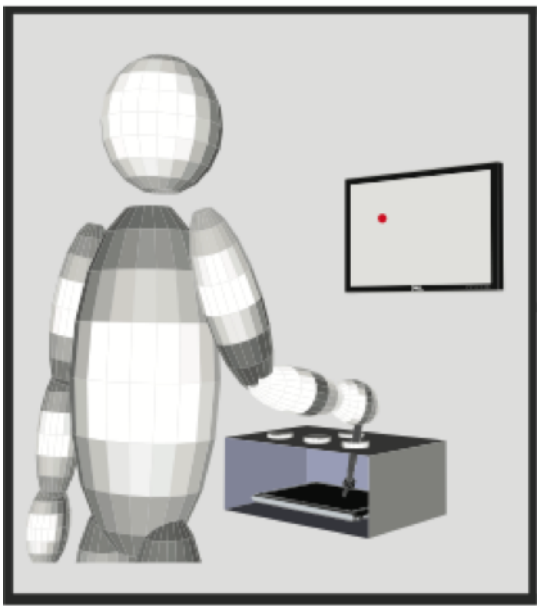


Fig. 3.2 Diagram of Laparoscopic Experimental Setup

3.1.1 ESOX Metrics

Post-performance processing of the output CSV files generated by ESOX allows for detailed evaluation of performance. From the relative cursor position, overall performance time, overall path length and time and distance between each coloured square can all easily be calculated. The ESOX task can only be completed in order, and consequently, deviation from the correct sequence will increase both time taken and distance travelled. Additionally, precise, controlled movements (desirable in laparoscopic surgery) reduce path length and time taken. Equal weighting is given to both time and path length and consequently a composite score of performance can be calculated:

$$ESOX\ Score\ (pxs) = distance(px) \times time(s) \tag{3.1}$$

Although often used in the analysis of sequence learning tasks [42], performance time alone is not utilised independently by ESOX. This is to try to better reflect the outcome of interest; the performance of laparoscopic surgery, where timely and *accurate* execution is necessary.

The above metric is appropriate when the position of coloured squares does not change between repeat trials. However, as described above, ESOX can randomly determine the position of squares, if requested to do so. In such an instance, the optimum distance between squares will differ with each repeat. In order to account for this vari-

ation, a modified compound score of performance can be employed:

$$\textit{Modified ESOX Score (}\boldsymbol{M}\textit{pxs)} = \frac{\textit{Actual Distance Travelled (px)}}{\textit{Optimum Distance (px)}} \times \textit{time(s)} \tag{3.2}$$

Where optimum distance is defined as the shortest possible distance between each individual layout of coloured squares.

Note: To allow comparisons to be drawn between different experimental setups and both the ESOX Score and *Modified* ESOX Score, a dimensionless Z-score is used at times in this thesis. This standardised score (z) of a raw score x is calculated thus:

$$z = \frac{x - \mu}{\sigma} \tag{3.3}$$

Where μ is the mean of the population and σ is the standard deviation of the population.

The position and time information recorded by the ESOX program also allows for the production of detailed topographical 'maps' of performance: Figure 3.3.

3.2 Premise of the ESOX Program

As described in Chapter 1, minimally invasive surgery can be thought of as a sequence learning task, requiring skilled, coordinated motor performance of sequential tasks. The components of the ESOX task have been designed to broadly mirror the concepts important in performing minimally invasive surgery. The sequence of component steps required to perform an operation is duplicated in ESOX, along with the identification of key components (coloured squares), which can be recognised through different modalities; colour and spatial location. Participants are required to demonstrate 'procedural knowledge' by performing ESOX, i.e. the sequence and location of squares and how to progress between them as well as developing 'technical ability', adapting to the visuomotor transformation inherent in minimally invasive surgery.

Spatial location and colour are used as discriminators in the ESOX program as they are fundamental components of visual perception [98]. Although most studies of visual attention have investigated the effect of location within the visual field [99], attention can be selectively deployed to visual features, such as colour and direction of motion, regardless of their locations [100–103]. Consequently, both spatial location and colour can be manipulated independently to assess their relative importance in learning and

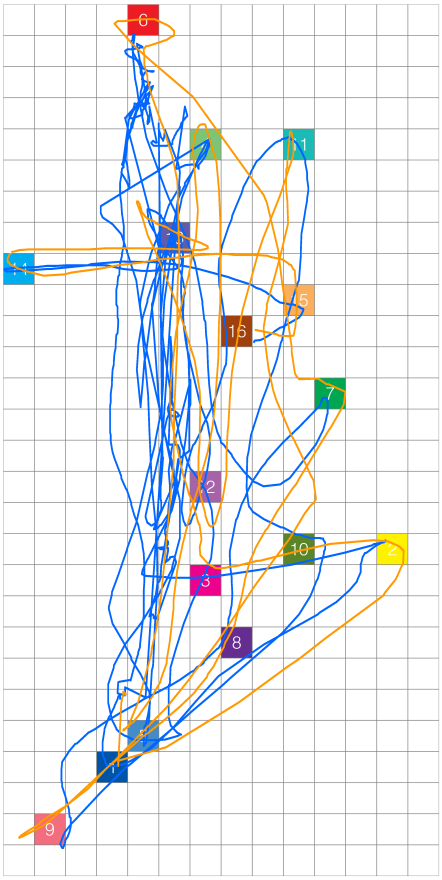


Fig. 3.3 Visual overlay of the path taken during two trials by the same participant. The blue line was produced at the first attempt, the orange line was produced on the eighth attempt.

performance facilitation.

Whilst the ESOX setup is removed from the reality of performing laparoscopic surgery, it has some important benefits; Participants can be rapidly taught to perform a task and, consequently, both learning and the effect of preoperative simulation can be quickly examined (*cf.* surgical training). Additionally, as the task is unique, the oft-found confound of differing levels of surgical experience is eliminated. Another significant advantage of the ESOX program is it allows the experimental conditions to be strictly controlled, something that is impossible in real-world laparoscopic surgery. Finally, ESOX can easily generate highly accurate metrics of performance to allow comparison between different performances. Again, this is something that is very difficult to achieve in real-world minimally-invasive surgery.

3.3 Experimental Designs

The ESOX program was used to conduct three randomised, controlled cross-over experiments - Experimentα (Chapters 4 and 5), Experimentβ (Chapters 6 and 7) and Experimentγ (Chapters 8 and 9). These experiments are discussed in detail in their relevant chapters but share some important characteristics.

Each experiment aims to examine the effect of preoperative simulation on learnt and consolidated (comparable to expert) performance. Previous work [104] has demonstrated that learning is a substantial confound to the effect of preoperative simulation; with repeated trials, novice participants improve their performance due to the effect of learning, which is difficult to parse from any effect of preoperative simulation. This confound can be removed by examining learnt performance, where any effect of learning is negligible. Whilst it is possible that preoperative simulation may play an important role in shortening a learning curve and hastening a plateau of performance, a systematic investigation of such is beyond the purview of this thesis.

Each experiment can be divided into two distinct phases - a learning phase and an assessment phase. During the learning phase, participants are taught to perform a specific laparoscopic sequence-learning task, through a series of complementary learning trials. The nature of each differs with each experiment, but they all share the same basic characteristics. Three learning trials are employed, the first two emphasising either motor performance or a more abstract conceptualisation of the task. Participants are randomly allocated to receive either one first, but all then progress to the final learning trial, which combines both aspects. In addition, all participants learnt the task on two separate days, permitting off-line processing and consolidation of the learnt skill(s).

Those participants who are able to demonstrate having learnt the task, are subsequently examined performing the laparoscopic sequence-learning task following a preoperative simulation; the assessment phase. How participants were deemed to have learnt is described below. Each assessment of preoperative simulation was temporally separated by a day, to reduce the effect of prior performance on the assessment. Participants were block randomised to perform each preoperative simulation in a random order, to reduce any effect of extended practice and avoid order effects.

3.3.1 Assessment of Learning

To avoid the confound of learning, participants had to demonstrate having learnt the specific task in each experiment during the learning trials. Participants who were unable to demonstrate having learnt were not included in the analysis of preoperative

simulation. Learning was assessed by reviewing repeated measures of performance. As

> "The benefits from practice follow a nonlinear function: improvement is rapid at first but decreases as the practitioner becomes more skilled..." [25]

simple non-linear functions have long been used to describe the effects of practice on multiple tasks [23], with the log-linear or power law for practice becoming ubiquitous [24]. The power law of practice is defined as [23]:

$$R = a + bN^{-\beta} \tag{3.4}$$

Where R is a random variable, a is the asymptote (reflecting the end of learning), b is the difference between initial and asymptotic performance, N is the amount of practice (measured as number of trials) and β is the rate of learning.

However, Heathcote, Brown and Mewhort [25] questioned the use of the power law as a descriptor for practice, particularly as they found the majority of research published since the introduction of the supposed power 'law' assumed a power function of learning, but did not examine the validity of other functions to describe the observed results. Heathcote, Brown and Mewhort [25] retrospectively reviewed the results of 40 sets of data, representing 7910 learning series from 475 subjects in 24 experiments taken from 13 published and 3 unpublished sources. The included data sets were drawn from memory search, counting, mental arithmetic, alphabetic arithmetic, visual search, motor learning, learning rules from examples and mental rotation paradigms and consequently correlate well with the type of learning necessary to complete the ESOX task. When reviewing the shape of the practice function, Heathcote, Brown and Mewhort [25] found the exponential function provided a better fit than the power function in 82.2% of cases. This result ranged from 64% to 93% for the individual included data sets, but in every case the authors found, at the 95% confidence interval, the exponential function to be a better descriptor of repeat performances. In addition, the exponential function accounted for more variance than the power function, with an R^2 of 0.498 for the exponential function $cf.$ 0.426 for the power function. On average, the exponential function provided a 17% [range 3.7% - 28.6%] improvement relative to the power function. Interestingly, it has been suggested by Rickard [105] that the reason the power law of practice fails to model learning behaviour is because subjects use a combination of memory and algorithmic-based process, which individually follow a power function, but do not once amalgamated. Heathcote, Brown and Mewhort [25] also explored this hypothesis, examining a subset of the included data sets which reported their processing strategies. They again found that the exponential function

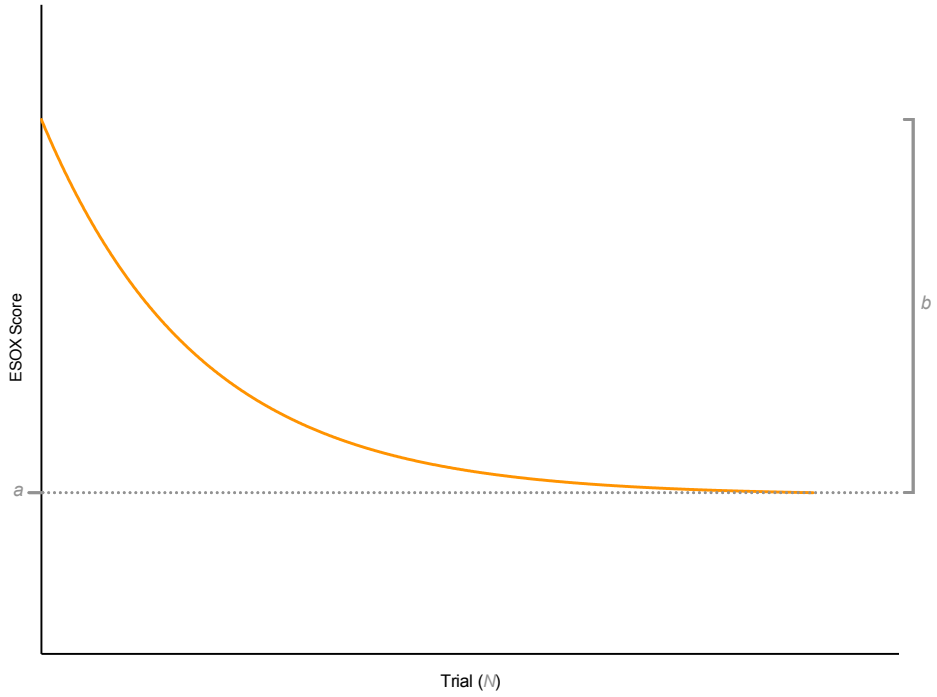


Fig. 3.4 Model of the Exponential Function of Learning.
a is the asymptote/plateau of performance, b is the difference between initial and asymptotic performance, N is the number of trials, and α is a rate constant of learning (not shown).

gave a better fit than the power function for both algorithmic (75.1% better) and memory (79.2% better) systems (overall 77.8% better), leading them to conclude that the better fit for the exponential function is not due to a mixture of component power functions. Overall, Heathcote, Brown and Mewhort [25] state that:

> "...practice produces a simple exponential improvement and a constant relative learning rate..."[2]

Consequently, it appears clear that learning is best modelled by the exponential function, defined as [23]:

$$R = a + be^{-\alpha N} \tag{3.5}$$

Where R is a random variable, a is the asymptote (reflecting the effective end of learning), b is the difference between initial and asymptotic performance, N is the amount of practice (measured as number of trials) and α is the rate of learning: Figure 3.4.

[2]The authors do add the caveat that some change in the relative learning rate may occur in early practice, but this is of little practical import for the assessment of learning required for this thesis; Establishing if a participant has successfully learnt (i.e. achieved plateau) was the aim, as opposed to the relative change in rate of performance improvement in the early stages of learning.

Throughout this thesis, the exponential function was used to model the learning of participants. Initially, the average performance across all participants was reviewed by averaging performance across individual curves. There have been concerns raised about arithmetically averaging learning curves across participants [106–109], as the parameters of the average function may not equal the average of the parameters of the component functions [110]:

> "...we might have an explanation of an average subject, but one that does not apply to any of the actual individuals making up the average." [111]

Understanding these concerns, there remains support for the use of averaged data, particularly as the use of averaged data is often useful at revealing general trends [25, 112].

After analysing the average performance across participants, individual participant's performance were reviewed. Individual learning curves were assessed by block averaging within curves, following the example of Brown and Heathcote [110]. Once again, averaging within curves is somewhat controversial, with Newell, Liu and Mayer-Kress [113] arguing that blocking data from groups of trials can modify or mask both transient and persistent changes. However, Brown and Heathcote show mathematically that block averaging N blocks of M trials results in a linear change of the scale parameter, but no distortion of shape of the exponential function (see Equations (3) and (4), page 16, Brown and Heathcote [110]). This mathematical proof was further consolidated by reanalysis of published data, which confirmed the Brown and Heathcote's findings. Consequently, Brown and Heathcote [110] conclude that block averaging can take advantage of the noise reduction inherent in arithmetic averaging, without introducing any averaging distortion to the exponential function.

Participants were deemed to have learnt if repeated performances during the learning trials following a decreasing exponential curve of performance with an arbitrary goodness of fit measurement of at least $R^2 = 0.5$ and reached a predicted plateau of performance. Plateau of performance (a) was calculated from the equation above (Equation 3.5, Figure 3.4) and mathematically represents the value of performance (ESOX Score) at infinite trials.

Chapter 4

Experimentα: Learning-Phase

Experimentα is the first of three randomised, controlled cross-over experiments, designed to explore the effects of preoperative simulation on subsequent laparoscopic performance under regulated, laboratory conditions. Experimentα can be can be divided into two stages - a learning phase and an assessment phase. The learning phase was designed to teach participants the experimental task, and the assessment phase was used to examine the effect of preoperative simulation on subsequent performance. This chapter discusses the initial learning phase.

4.1 Experimental Task α

The experimental task is a sequence-learning task that is designed to contain both spatial, kinematic and colour components. Participants are shown a sequence of numbered, coloured squares on a 14 by 28 array and are asked to move a laparoscopic stylus to each square, in sequence: Figure 3.1. Though participants are initially shown the sequence order, once the program is started by pressing on the tablet, the numbers are removed, necessitating the participants to remember the sequence order of coloured squares. To enhance learning, once the correct square is reached, it is no longer visible, reinforcing to participants that they have reached the correct sqaure.

The Experimentα task was designed to give equal weighting to both spatial and colour information - the task could be successfully completed by utilising just one of these descriptors, or a combination of both.

4.2 Methods

The aim of this phase of the experiment was to teach the participants the Experimentα task. Though this task was designed to contain spatial, kinematic and colour components, previous work has demonstrated that under such a system, participants tend to only learn spatial information and are unable to continue to follow the sequence if a spatial transformation is applied [114]. Consequently, to try to encourage participants to learn spatial, kinematic and colour information, participants progressed through three stages of training:

1. Learning Kinematic and Spatial Information (αLearn:K&S): in this task, the squares were distributed in the array in the same manner as the Experimentα task, but no colour information was provided (all squares were coloured grey): Figure 4.1. This task was performed using a laparoscopic stylus in a box trainer.

2. Learning Colour Information (αLearn:C): in this task, all squares were distributed evenly around a centre point. Colour information was provided and maintained throughout each repetition (i.e. Square 1 was always blue), but the position of each square in each repetition was randomly determined by applying a Fisher-Yates shuffle [97] before each array was built: Figure 4.2a, Figure 4.2b. To differentiate between spatial and colour information, and to remove the inherent kinematic transformation of laparoscopic surgery, this task was not performed in a laparoscopic box trainer but instead on a separate tablet computer using a normal pen-like stylus.

3. Learning Colour, Kinematic and Spatial Information (αLearn:C,K&S): This trial combined both of the above tasks; the positioning of squares was maintained from Task 1 while the colouring of squares was as per Task 2. Consequently, participants could use spatial information, colour information, or combination of both to complete the task: Figure 3.1. This task was performed using a laparoscopic stylus in a box trainer. This task has the same configuration as that used during assessment.

Participants were randomised to receive either Task 1 then Task 2, or Task 2 and then Task 1. Subsequently, all participants completed Task 3 (αLearn:C,K&S). This randomisation was performed to ensure that participants did not gain a significant advantage by performing the tasks in a certain order. To further try and enhance learning, after each trial (for all tasks), participants were automatically given feedback about their performance from the ESOX program in the form of time taken (seconds

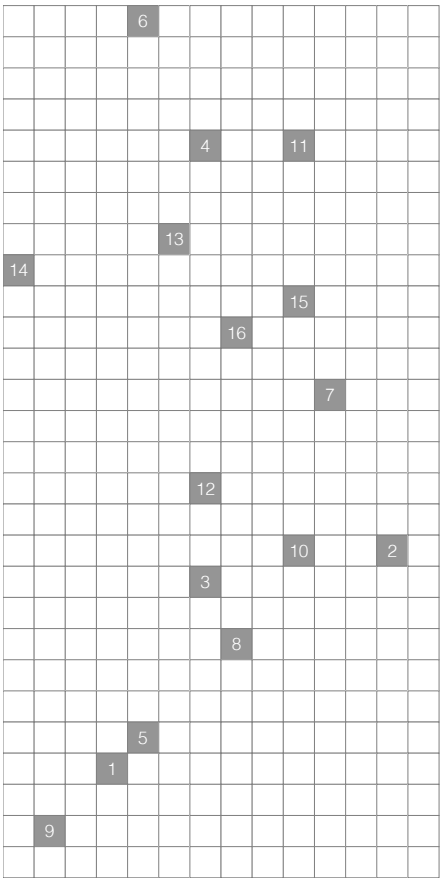
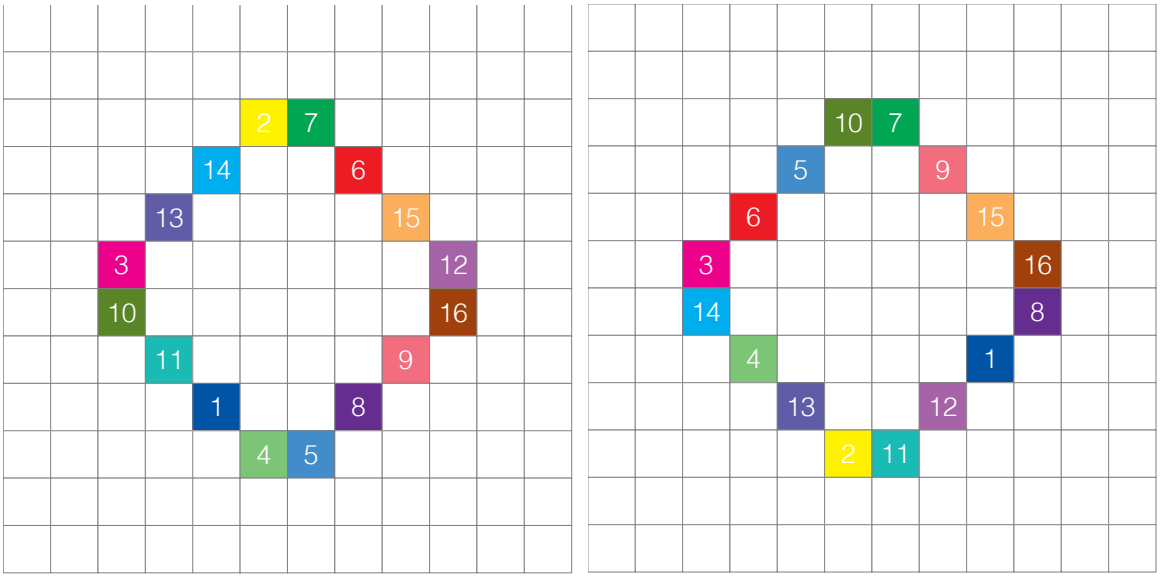


Fig. 4.1 Array for Learning Kinematic and Spatial Information (αLearn:K&S)

(s)) and distance travelled (pixels (px)) while completing the task. Such feedback has been demonstrated to improve student learning of technical surgical skills [115–118]. This information was displayed on the screen after each task was completed; Figure 4.3.

All participants were trained to complete Tasks 1, 2 and 3 on two separate occasions. These training sessions were temporally limited to a maximum of one hour. As participants had to learn three tasks, they were encouraged to spend approximately 20 minutes on each. However, this timing was not fixed, as pilot data suggested spatial information was easier to learn than colour information in the above experimental paradigm. In addition, previous work has shown that people learn at different rates and can reach differing plateaus of performance [119–121]. If participants felt they had learnt the tasks prior to one hour, the session could be terminated early by the participant. Sessions were performed on separate days to enable off-line processing and consolidations of the tasks [57–60].



(a) Example 1 **(b)** Example 2

Fig. 4.2 Arrays for Learning Colour Information (αLearn:C)

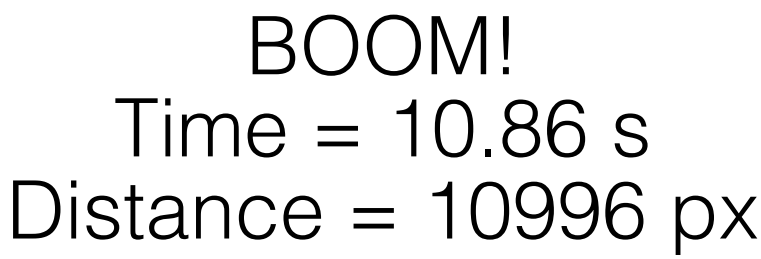


Fig. 4.3 Feedback given to participants at the end of the ESOX task.

4.3 Analysis

Performance was assessed using the metrics discussed in §3.1.1 and learning was analysed using the methods detailed in §3.3.1. The demographics of study participants were analysed and are reported as medians with ranges. Performance at preoperative simulation (discussed in Chapter 5) was analysed using a repeated measures analysis of variance (ANOVA), comparing performance following the control condition to other preoperative simulation routines. These results were graphed, with within-subject error bars being computed using a modification of Loftus and Masson's method [122] as proposed by Bakeman and McArthur [123], Cousineau [124], Morey [125], and Morrison and Weaver [126]. The effect of learning trial order on performance during the learning trials and performance following preoperative simulation was analysed using an independent samples t-test and an analysis of covariance (ANCOVA), respectively.

Statistical analysis was performed using Prism, Version 6.0C (GraphPad Software, Inc. La Jolla, CA, USA), IBM SPSS Statistics for Macintosh, Version 21.0. (IBM Corp. Armonk, NY, USA) and G*Power: Statistical Power Analyses for Mac, Version 3.1.9.2 [127].

4.3.1 Sample-size Calculation

Data were analysed using a repeated measures analysis of variance (ANOVA). To account for a moderate effect size ($f = 0.25$), the most conservative estimate based on pilot data taken from previous experimentation [104] and the development of the ESOX program, *a priori* power calculations showed 25 participants (in total) would be required to achieve an α error probability of 0.05 and power (1-β error) of 0.8. To ensure this sample size was achieved (accounting for potential drop outs), we aimed to recruit 40 participants.

4.3.2 Participants

Participants were recruited from the Participant Pool Scheme employed by the School of Psychology, University of Leeds. Such participants were chosen as being representative of the surgically-naive population, being drawn from students and staff at the University of Leeds. Participants with motor or cognitive impairment, impaired colour vision, or previous surgical experience were excluded from the study. Participants were reimbursed £30 for their time.

4.3.3 Randomisation and Blinding

The order participants performed the preoperative simulation routines was determined by a stratified, permuted block randomisation sequence with a block size of 5 and an allocation ratio of 1:1. All assessment metrics were calculated by computer and consequently blind.

4.3.4 Ethical Approval

Experimentα received ethical approval from the School of Psychology, The University of Leeds (ref: 15-0038, 04-Feb-2015 and 15-0199, 06-Aug-2015).

4.4 Results

Forty participants were recruited, with thirty-eight undertaking all stages of the study (26 female : 12 male, average age 28 years [range 19 - 51], 32 right-hand dominant: 6 left-had dominant).

4.4.1 Averaged Learning Results

Although interpreted with some caution (see § 3.3.1), review of the average learning curves across participants revealed two interesting findings.

All conditions (αLearn:K&S, αLearn:C, αLearn:C,K&S) demonstrated a consistent pattern across the average of participants of exponential decay - rapid early improvement, followed by a phase of slower incremental performance improvement and eventual plateau: Figure 4.4. The exponential function models the averaged results very well, as shown by high coefficients of determination (R^2): Table 4.1. It is also worth noting that the power law of practice (log-linear) did not model the observed results as well as the exponential function overall: Table 4.2.

Secondly, review of the raw averaged results of the αLearn:K&S and αLearn:C,K&S trials demonstrate both closely follow an exponential decay pattern (and as evidenced by the high coefficients of determination in Table 4.1), but when compared directly, αLearn:C,K&S shows a *comparatively* flatter curve of performance improvement: Figure 4.5.

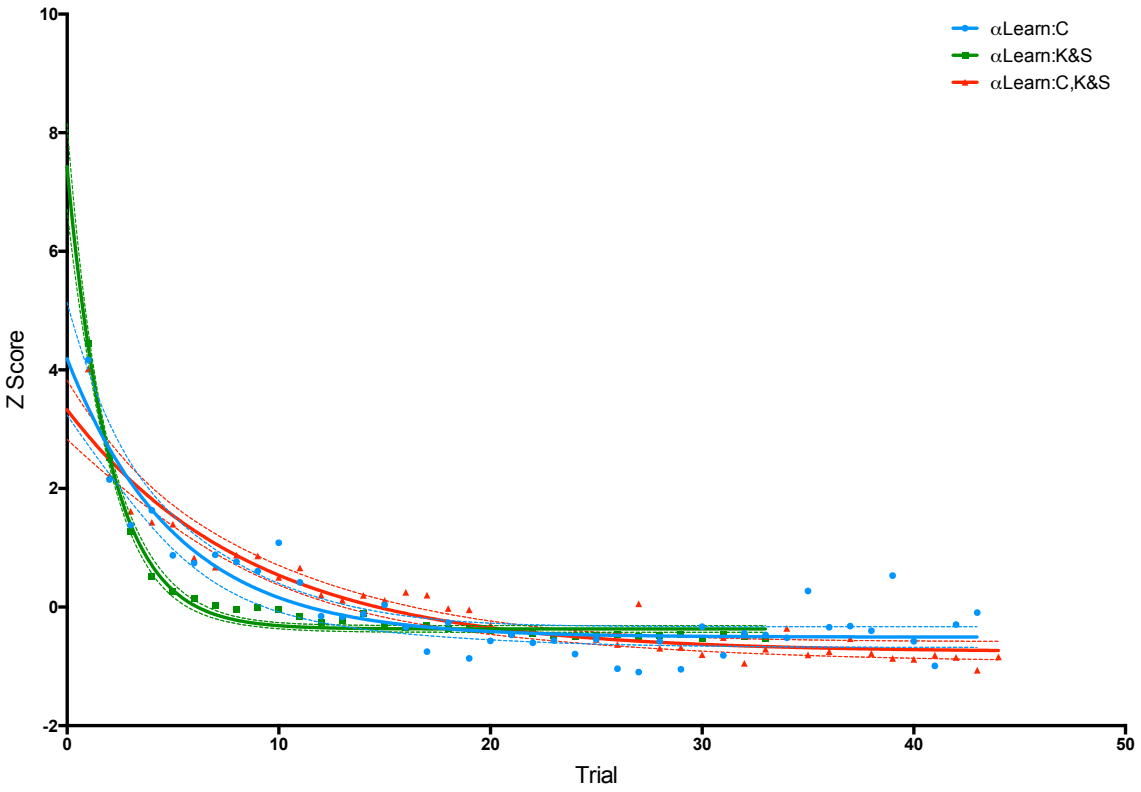


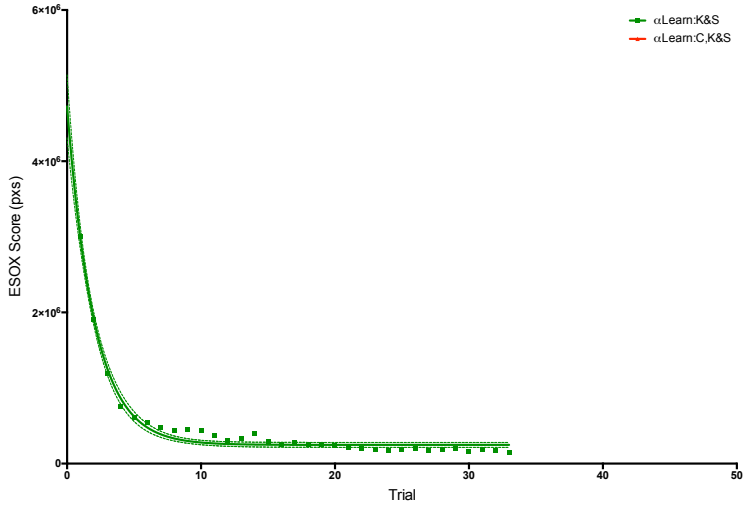
Fig. 4.4 Average Performance during Experimentα Learning Trials.
Learning Colour Information trials (αLearn:C) are represented in blue, Learning Kinematic and Spatial Information trials (αLearn:K&S) are shown in green and Learning Colour, Kinematic and Spatial Information trials (αLearn:C,K&S) in red. The dotted lines represent the 95% confidence interval (CI) of the mean.
Note: Results are shown as dimensionless Z-scores to allow comparison between different trials.

Table 4.1 Coefficients of determination (R^2) for the averaged results of Experimentα Learning-Phase

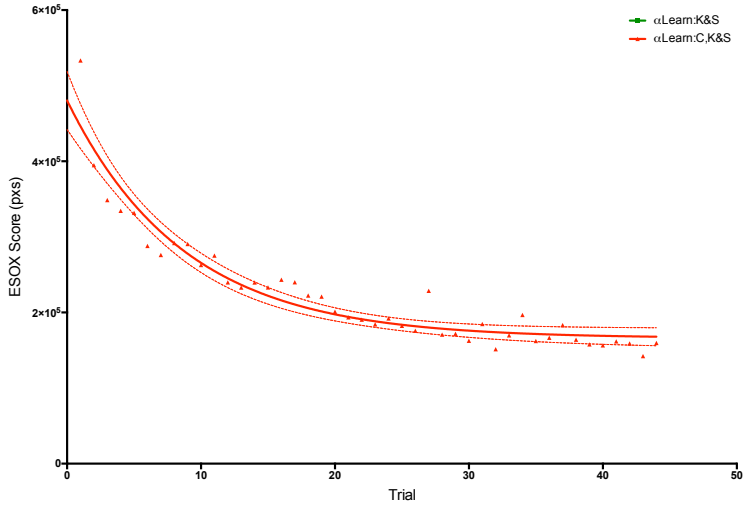
Learning Trial	R^2
αLearn:C	0.8361
αLearn:K&S	0.9820
αLearn:C,K&S	0.9157

Table 4.2 Coefficients of determination (R^2) for the averaged results of ESOXα Learning Phase (Log-linear Function)

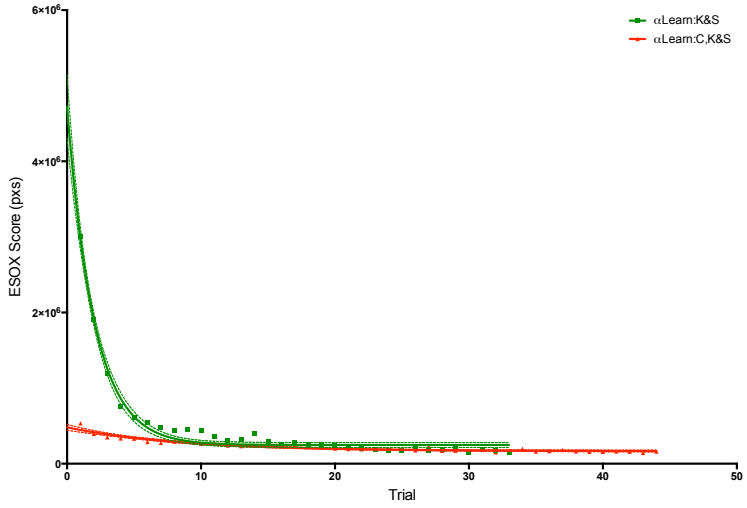
Learning Trial	R^2
αLearn:C	0.7435
αLearn:K&S	0.7683
αLearn:C,K&S	0.9458



(a) Average Performance at αLearn:K&S



(b) Average Performance at αLearn:C,K&S



(c) Graph **(a)** and **(b)** displayed on a shared y-axis

Fig. 4.5 Average Performance during αLearn:K&S and αLearn:C,K&S.
Note: A difference in y-axis scale in the first two graphs.

4.4.2 Individual Learning Results

Blocked individual learning results were compared to an exponential function of learning, as described in §3.3.1. The coefficients of determination, predicted plateaus of performance and the smallest (best) result achieved during the Learning Phase are reported in Tables D.1, D.2 and D.3 in Appendix D.

The order participants undertook the learning trials did not have an effect on their performance at plateau: an independent samples t-test indicated no difference in performance at plateau for those participants who progressed through αLearn:K&S, αLearn:C and then αLearn:C,K&S learning tasks ($\bar{x} = -0.173925$, $\sigma = 1.008262$) when compared to those who performed αLearn:C, αLearn:K&S and finally αLearn:C,K&S ($\bar{x} = -0.190484$, $\sigma = 0.713461$), $t(36) = 0.058$, $p = 0.954$, $d = 0.02$.

4.5 Discussion

The above results demonstrate the majority of participants were able to satisfactorily learn the laparoscopic sequence learning task, though a number of participants failed to reach a plateau of performance at one of the three learning trials. Participants appear to have found learning the kinematic and spatial (αLearn:K&S) task (32 successful) easier than learning colour, kinematic and spatial (αLearn:C,K&S) task (20 successful), and both easier than colour (αLearn:C) task (16 successful). This seems somewhat counterintuitive as the αLearn:C,K&S trial contains the kinematic and spatial information from the αLearn:K&S trial and the colour information from the αLearn:C trial. Consequently, it seems reasonable to assume that information from the first two trials would inform performance at the αLearn:C,K&S trial. There are a number of potential, complementary explanations for this apparent paradox:

1. Participants only performed the αLearn:C,K&S trial *after* completing the other two trials. Consequently, those participants that were unable to reach a plateau of performance may have been trying to use a different strategy to complete the αLearn:C,K&S trial than they had previously successfully employed during αLearn:K&S and αLearn:C. In such a scenario, the preceding trials may have acted as confounds to each other, hampering performance during the αLearn:C,K&S trial for some participants.

2. Individual participant's performance at plateau was significantly better for the αLearn:C,K&S trial than the αLearn:K&S trial: a paired-samples t-test showed that ESOX scores were significantly lower at plateau during αLearn:C,K&S ($\bar{x} =$

234857pxs, σ = 126014pxs) than during αLearn:K&S ($\bar{x}$ = 379316pxs, σ = 351189pxs), $t(16) = 2.46$, $p = 0.025$, $d = 0.55$. Participants who did not achieve a plateau of performance during the αLearn:C,K&S trial may have terminated their repetitions when they reached a similar performance level to the αLearn:K&S trial, while others continued to an improved plateau.

3. As shown in Figure 4.5, participants demonstrated a substantially flatter curve during αLearn:C,K&S compared to αLearn:K&S, on average. Accordingly, as αLearn:C,K&S can be completed by utilising kinematic and spatial information alone, it may be that some participants *only* utilised kinematic and spatial information to complete αLearn:C,K&S and consequently had already reached their plateau of performance during αLearn:K&S. Similarly, very poor initial performance with substantial subsequent improvement may appear to exhibit 'better' learning than mediocre initial performance which improves to a lower plateau of performance under the assessment of learning framework utilised during this experiment (see §3.3.1).

Participants seemed to have the most difficulty in learning the αLearn:C trial. All trials were designed to be challenging, but achievable, and were extensively trialled during their development utilising both surgeons and surgically-naive volunteers, prior to experimentation. It has been argued that spatial location is a 'special' feature in selecting information and assumes priority over other modalities [128, 129] and is the default of selective attention [130, 131]. Consequently, participants may have found it difficult to utilise other information (i.e. colour) during the αLearn:C trial, particularly as some may have been primed to use spatial location information during prior performance of αLearn:K&S. There is no consensus regarding the primacy of location information, but it does appear to be activated before other feature-based mechanisms [99, 132]. The time pressure exerted on participants (they were asked to complete the task as fast as they were able), could have prompted certain participants to utilise the first discriminator available to them - spatial location - which was unhelpful in completing αLearn:C.

Alternatively, it may be that participants simply did not find the αLearn:C trial as interesting or captivating as the other trials and consequently invested less time and effort in this trial. Anecdotally, as participants were given feedback about their performance at the end of all trials (time taken and distance travelled), they tended to become competitive and try to beat their previous performance during the αLearn:C,K&S and αLearn:K&S trials. This did not seem to happen during the αLearn:C trial. Again anecdotally, participants tended to become more frustrated during the αLearn:C

trial, which may have hampered their progression. Removing feedback from all trials may adjust for this in further experimentation (though participants would still be able to infer their relative performance without explicit feedback). Additionally, both αLearn:C,K&S and αLearn:K&S were conducted using a laparoscopic stylus in a laparoscopic box trainer. The αLearn:C trial was conduced on a tablet PC. The included participants were all interested in performing simulated laparoscopic surgery (as they self-selected themselves for the study). The fact that the αLearn:C trial seemed to have less to do with laparoscopic surgery may have decreased some participants interest (and consequently their performance).

All participants were able to demonstrate having learnt at least one of the learning trials, but only four participants (Participant No. 7, 25, 29 and 37) were able to demonstrate having learnt all three learning trials. This does suggest that the Experimentα tasks are not trivial, but require significant motoric control and cognitive performance. That the Experimentα tasks require high levels of cognitive and motor performance implies they are good surrogates for laparoscopic surgery (another task that cannot be considered trivial).

Chapter 5

Experimentα: Assessment of Preoperative Simulation

Those participants that were able to demonstrate having successfully learnt the αLearn tasks were assessed following a preoperative simulation, to examine the effect of preparation on learnt performance. As discussed in Chapters 1 and 2, there is theoretical and literature evidence to support the notion that preparation can improve performance. This Chapter aims to delineate the effective components (if any) of a preparatory routine for the experimental task.

5.1 Methods

Once they had successfully progressed through the Learning-Phase, participants were asked to perform five preoperative simulation exercises (including one control), one per day, in a randomly determined order, prior to performing an assessment task. This phase of the experiment took place in the week following the learning phase. The assessment task was very similar to the αLearn:C,K&S task described in §4.2, as the same spatial and colour information was maintained, but it did not give the participants any order information (there were no numbers in the squares): Figure 5.1. Each assessment was temporally separated by one day.

Participants were asked to perform one of each of the following preoperative simulation routines, in a randomly determined order, prior to performing the assessment task.

1. Kinematic and Spatial Preoperative Simulation (αPS:K&S): Participants were asked to repeat Learning Task 1 (αLearn:K&S). This provided participants with spatial and kinematic, but no colour information.

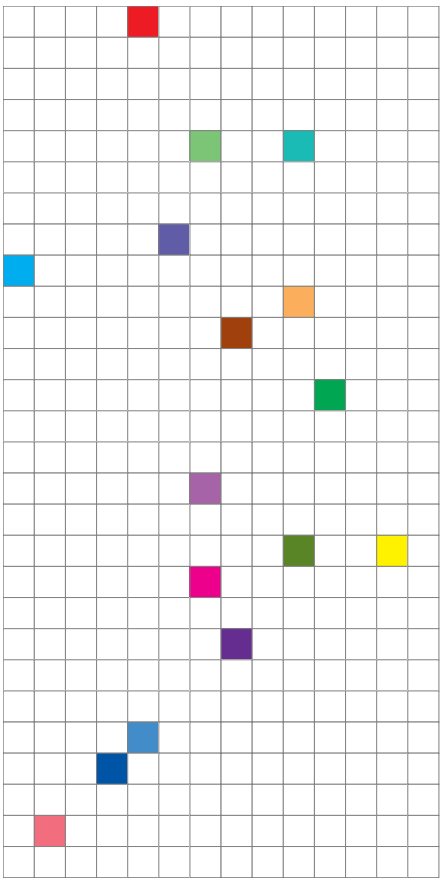


Fig. 5.1 Array for Assessment (without sequence information)

2. Colour Preoperative Simulation (αPS:C): Participants were asked to repeat Learning Task 2 (αLearn:C). This provided the participants with colour, but no spatial or kinematic information relating to the assessment task.

3. Observe Performance of Task (αPS:O): Participants were asked to view a recording of ESOXα being performed. This provided participants with both spatial and colour information, but without direct physical involvement from the participant (no direct kinematic information). The recording was shown on the laparoscopic screen used to perform the ESOXα task.

4. Colour, Kinematic and Spatial Preoperative Simulation (αPS:C,K&S): Participants were asked to perform the sequence-learning task (αLearn:C,K&S), immediately prior to performing the assessment task (specific warm-up).

5. Control Group: Participants were asked to play Tetris®. This was chosen to occupy participants during the preoperative simulation period, without giving them any specific information about the upcoming assessment task. Although

not specific to the experimental task, Tetris®involves visual identification based on spatial and colour information, a kinematic transformation between physical action and outcome and brightly coloured squares displayed on a computer monitor. Consequently, while not directly related to the Experimentα task, it could be considered a general preparation of the visual-motor system, as employed by some studies in Chapter 2 [74, 81]. This preoperative simulation routine was performed utilising the monitor on the laparoscopic stack, but was completed using a keyboard.

Each preoperative simulation trial was limited to three minutes. This was controlled by computer script in all cases, aside from the control group which was timed by a supervising investigator. Participants could repeat the preoperative simulation routine as many times as was their preference during this time. This time restraint, chosen from pilot data, was employed to ensure parity between different participants and preoperative simulation routines. To further ensure this, each participant was given the same written instruction before each preoperative simulation and assessment: Appendix E.

The above preoperative simulation routines were designed to offer the participants varying amounts of colour, spatial and kinematic information. Repeated performance of the assessment task (αLearn:C,K&S) was included as the theoretical 'ideal' preoperative simulation.

5.2 Results

Three participants (Participant No. 6, 16 and 33) were excluded from analysis of preoperative simulation because their performance during assessment were more than 3 standard deviations away from the mean of other participants' performance (3.48, 4.98 and 10.3, respectively). They were consequently deemed to be statistical outliers and thus excluded. Unfortunately, data for one participant's performance following one preoperative simulation was lost due to a technical error. Consequently, this participant (Participant No. 7) also had to be excluded from further analysis. The following results are therefore based on 34 participants' results.

Analysis of performance following preoperative simulation was assessed using a repeated measures ANOVA. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2 = 40.6$, $p < 0.001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.677$). A main effect of preoperative simulation on performance, $F(2.71,\ 89.32) = 3.184$, $p = 0.0321$, $\eta_p^2 = 0.088$, was demonstrated. Post-hoc analyses using Fisher's LSD revealed this was driven by

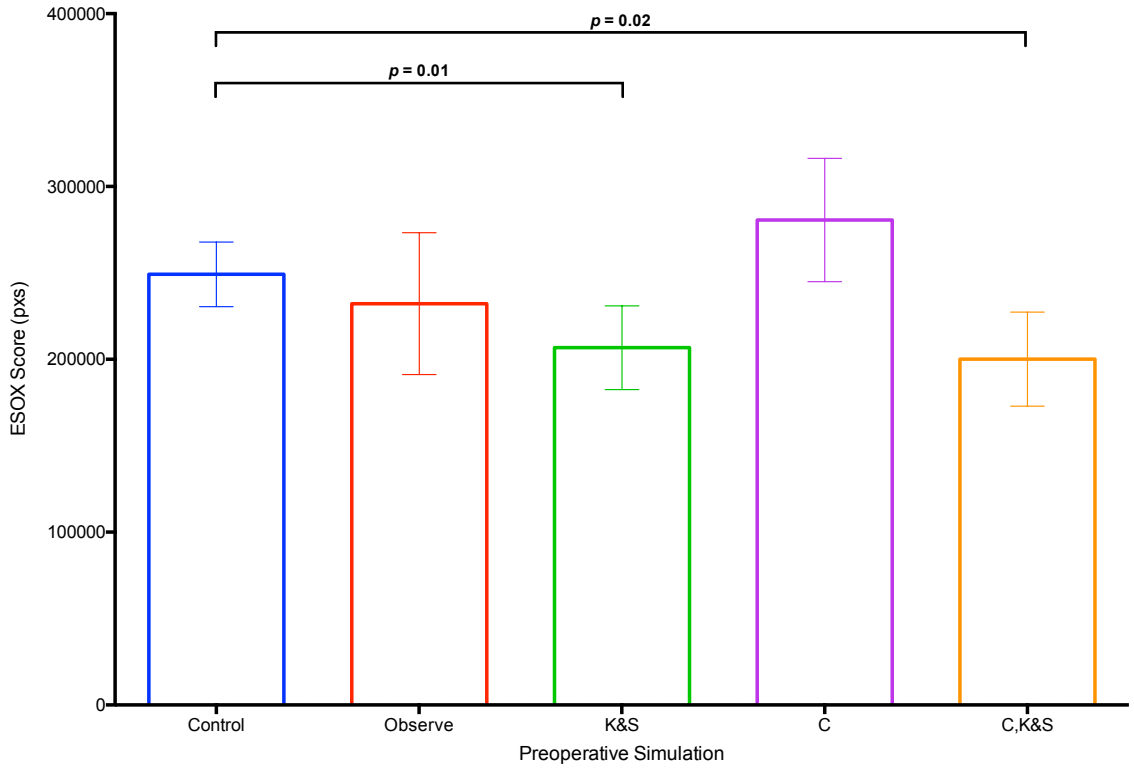


Fig. 5.2 Performance following Preoperative Simulation during the Experimentα. Significant within-subject effects are highlighted. The error bars represent the 95% CI of the mean. **Note**: Smaller scores represent better performance.

significantly better performance following a kinematic and spatial (K&S) preoperative simulation when compared to control ($p = 0.0108$) and better performance after a colour, kinematic and spatial (C,K&S) preoperative simulation when compared to control ($p = 0.0211$). These results are summarised in Figure 5.2.

The order in which participants performed the learning trials did not affect their performance at assessment: an ANCOVA [between-subjects factor: preoperative simulation (Control, Observe, K&S, C and C,K&S); covariate: learning order] revealed no interaction between preoperative simulation and learning order, $F(1, 32) = 0.748$, $p = 0.513254$, $\eta_p^2 = 0.023$.

5.3 Discussion

A repeated-measures ANOVA of the four different preoperative simulation routines (and one control) demonstrated significant differences between the groups. Post-hoc analysis revealed that this difference was driven by changes across the control condition and the K&S preoperative simulation and between the control and αPS:C,K&S.

In addition, αPS:C seemed to *diminish* performance (although it did not reach significance when compared to the control condition). Indeed, there is a significant difference between performance following the αPS:C and both αPS:K&S ($p = 0.007$) and αPS:C,K&S ($p = 0.012$). These results have some interesting implications for surgery:

1. Preoperative simulation routines may have an effect on subsequent performance. These results suggest that not only could an effective preoperative routine enhance performance, but that an 'incorrect' routine may adversely affect performance (and potentially, by extension, patient outcomes).

2. Preoperative simulation routines may have to be specific to the information that is important to the operator, which is likely to reflect how the operator learnt the procedure. In this experiment participants were assessed performing the Experimentα task, which contains both spatial and colour information. There is robust evidence [133–135] that such information is retained as a bound representation in memory:

   > "... we do not remember a jumble of different kinds of information or features. We do not remember blue, brown, pen, table, but rather a blue pen on a brown table." [134].

   There is equally strong evidence that the binding of this information is not symmetrical, i.e. not all features of an object are equally important [136, 137]. That a spatial preoperative simulation routine was beneficial, whereas a colour routine was detrimental, suggests that spatial information is the dominant aspect of the bound information. Other studies in the literature have reported similar findings [134], as well as work within our group [114].

3. The preoperative simulations that provided participants with *relevant* information were effective. Participants utilised spatial information during completion of the Experimentα task, as shown during the learning trials and as discussed above. Subsequently, those PS that provided participants with spatial and kinematic information (αPS:K&S and αPS:C,K&S) were effective preoperative simulations. As both of these PS were useful, and produced similar results, it may well be that the colour information provided by αPS:C,K&S was not utilised by participants, but instead participants were able to disregard this information and extract the useful spatial information from αPS:C,K&S - it would appear the addition of colour in αPS:C,K&S (*cf.* αPS:K&S) did not help participants. αPS:O also provided participants with spatial information, but did not help subsequent

performance. αPS:O did not allow participants to prepare for the kinematic transformation involved in the assessment task, which may explain this difference, and/or, active use of the motor system may be required for an effective spatial preoperative simulation.

Chapter 6

Experimentβ: Learning Phase

The results of Experimentα demonstrated that a greater number of participants were able to learn the αLearn:K&S task than the other learning tasks, and that participants reached a plateau of performance most quickly when performing αLearn:K&S (see Figure 4.4). In addition, preoperative simulations that provided spatial and kinematic information were effective (αPS:K&S, αPS:C,K&S). Conversely, participants had most difficulty in learning the αLearn:C task (as evidenced by the greater number of participants who did not reach plateau) and αPS:C appeared to, if anything, diminish performance. The Experimentα task was designed to give equal weighting to both spatial and colour information and could be successfully completed by utilising just one modality. Under such a paradigm, it would appear that participants will preferentially utilise spatial information when completing a sequence learning task that contains bound spatial and colour information.

As discussed in Chapter 3, the spatial and colour components of the ESOX task were utilised as they are fundamental components of colour vision [98], but also to try and represent the different fundamental aspects of skill/knowledge required to perform minimally-invasive surgery. Crudely, one might see the spatial information in this task as analogous to the physical movements required to perform surgery, while colour information maps onto the more abstract academic knowledge necessary. Experimentα indicates participants can learn the physical movements necessary to complete an ESOX task, and will preferentially utilise spatial information when spatial or colour information could be used. However, surgeons are unable to utilise spatial information alone when performing minimally-invasive surgery in the real-world. Consequently, Experimentβ was performed in order to explore the effects of preoperative simulation in a colour-dependent paradigm. This experiment was designed to assess if the same forms of preoperative simulation that were beneficial in Experimentα also con-

ferred an advantage when different descriptors of information were of primary import. Experimentβ is the second of three randomised, controlled cross-over experiments.

6.1 Experimental Task β

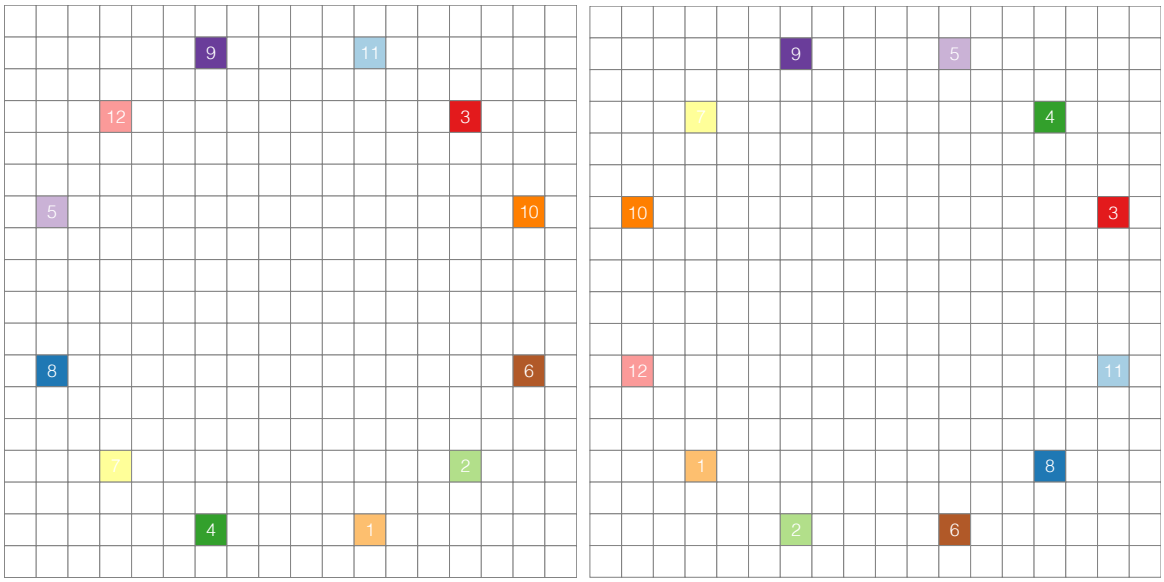
The Experimentβ task shares a number of characteristics with the Experimentα task. It remains a sequence-learning task that contains spatial, kinematic and colour components. However, in contrast to Experimentα, Experimentβ was designed to give predominance to colour information. Participants were shown a sequence of numbered, coloured squares on an 18 by 18 array and asked to move a laparoscopic stylus to each square, in order. Though participants are initially shown the sequence order, once the program is started by the participant, the numbers are removed, necessitating the participants to remember the sequence order of coloured squares. To aid learning, once the correct square was reached, it 'disappears'. Colour information was maintained throughout each repetition, but the position of squares within the circular design was randomised each time the program was reloaded by applying a Fisher-Yates shuffle [97] before each array was built: Figures 6.1a, 6.1b.

Two further changes to the experimental task were made (*cf.* Experimentα). The sequence of squares was empirically reduced from 16 to 12 to try and promote complete learning of the sequence by all participants. Additionally, in view of verbal feedback from participants in Experimentα, the colour of squares was modified (see Figures 6.1a, 6.1b) to increase discrimination between different squares. ColorBrewer [138], a cartographical tool for producing maximally diverging colour schemes and consequently legible maps, was employed to determine the colours.

6.2 Methods

As in Experimentα, participants progressed through three stages of training:

1. Learning Kinematic Information (βLearn:K): in this task, the squares were distributed in the array in the same manner as the Experimentβ task, but no colour information was provided (all squares were coloured grey). Each time the task was repeated, a Fisher-Yates shuffle [97] randomly determined the location of squares within the circular arrangement: Figures 6.2a, 6.2b. As no colour information was provided, upon starting the trial, the numbers within the squares remained (they did not disappear). This task was performed using a laparoscopic



(a) Example 1 **(b)** Example 2

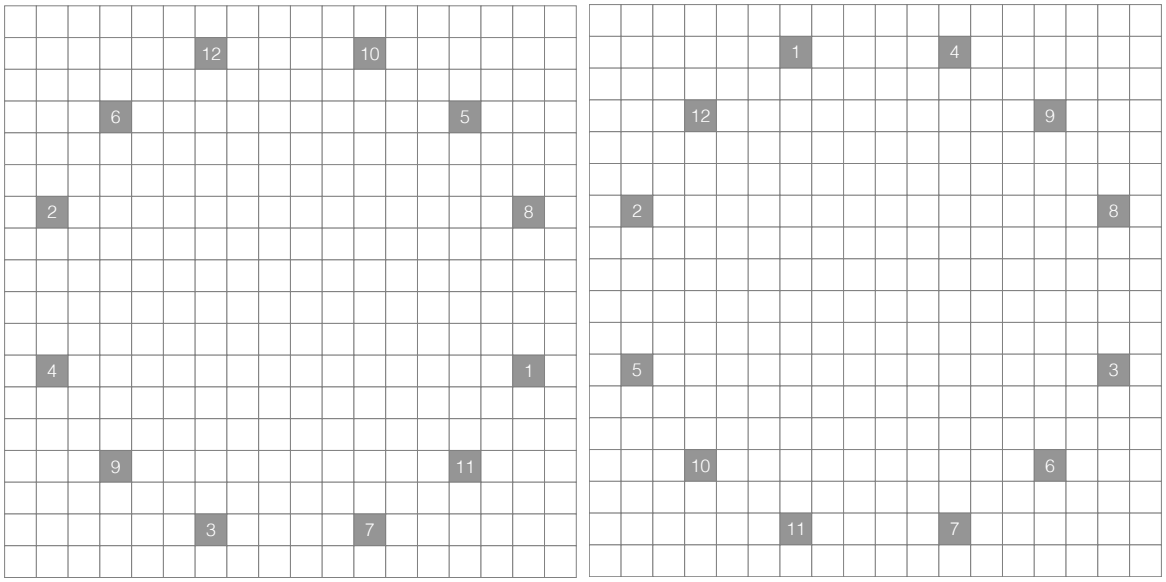
Fig. 6.1 Examples of Experimentβ output (βLearn:C&K)

stylus in a box trainer (see Figure 3.2) to mirror the kinematic transformation that occurs during the assessment task.

2. Learning Colour Information (βLearn:C): in this task, colour information was provided and maintained throughout each repetition (i.e. Square 1 was always peach-coloured), but the position of each square within the circular arrangement was randomly determined by applying a Fisher-Yates shuffle [97] before each array was built: Figures 6.3a, 6.3b. This task was not performed in a laparoscopic box trainer but instead on a separate tablet computer using a normal pen-like stylus, thereby providing no information regarding the kinematic transformation inherent in (simulated) laparoscopic surgery.

3. Learning Colour and Kinematic Information (βLearn:C&K): This trial combined both of the above tasks; the distribution of squares was maintained from Task 1 while the colouring of squares was as per Task 2. Once again, each time the task was performed, a Fisher-Yates [97] determined the position of coloured squares within the circular arrangement: Figure 6.1a, 6.1b. This task was performed using a laparoscopic stylus in a box trainer and had the same configuration as that used during assessment.

Mirroring Experimentα, participants were randomised to receive either Task 1 then Task 2, or Task 2 and then Task 1. Subsequently, all participants completed Task 3 (βLearn:C&K). To further try and enhance learning, after each trial (for all tasks),



(a) Example 1 **(b)** Example 2

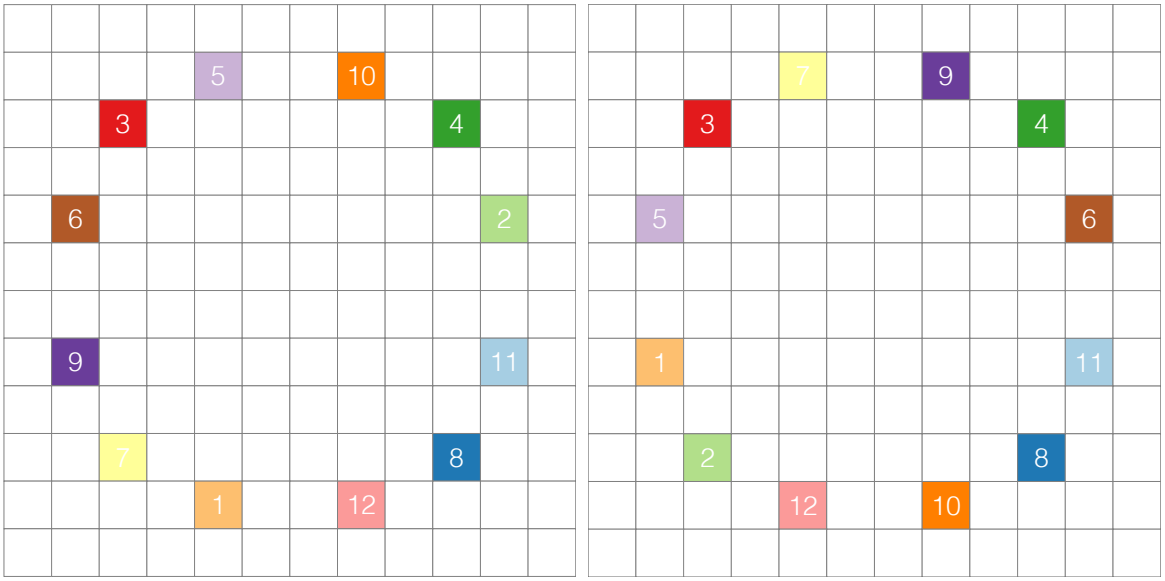
Fig. 6.2 Array for Learning Kinematic Information (βLearn:K)

participants were automatically given feedback about their performance from the ESOX program in the form of time taken and distance travelled in completing the task (Figure 4.3). Such feedback has been demonstrated to improve student learning of technical surgical skills [115–118].

All participants were trained to perform all three tasks on two separate occasions. These training sessions were temporally limited to a maximum of one hour. As participants had to learn three tasks, they were encouraged to spend approximately 20 minutes on each. However, this timing was not fixed, as the results of ESOXα demonstrated that spatial information was easier to learn than colour information in the above experimental paradigm. Equally, participants were shown to learn at different rates and reach differing plateaus of performance, a finding supported by other studies [119–121]. If participants felt they had learnt the tasks prior to one hour, the session could be terminated early by the participant. Sessions were performed on separate days to enable off-line processing and consolidations of the tasks [57–60].

6.3 Analysis

Performance was assessed using the metrics discussed in §3.1.1 and learning was analysed using the methods detailed in §3.3.1. The demographics of study participants were analysed and are reported as medians with ranges. Performance at preoperative simulation was analysed using a repeated measures analysis of variance (ANOVA), com-



(a) Example 1 **(b)** Example 2

Fig. 6.3 Array for Learning Colour Information (βLearn:C)

paring performance following the control condition to other preoperative simulation routines. These results were graphed, with within-subject error bars being computed using a modification of Loftus and Masson's method [122] as proposed by Bakeman and McArthur [123], Cousineau [124], Morey [125], and Morrison and Weaver [126]. The effect of learning trial order on performance during the learning trials and performance following preoperative simulation was analysed using an independent samples t-test and an analysis of covariance (ANCOVA), respectively. Models of learning were compared using Akaike's Information Criterion (AIC_c, described in detail in §6.4.1).

Statistical analysis was performed using Prism, Version 6.0C (GraphPad Software, Inc. La Jolla, CA, USA), IBM SPSS Statistics for Macintosh, Version 21.0. (IBM Corp. Armonk, NY, USA) and G*Power: Statistical Power Analyses for Mac, Version 3.1.9.2 [127].

6.3.1 Sample-size Calculation

Data were analysed using a repeated measures ANOVA. A similar effect size to Experimentα was assumed ($f = 0.311$) and consequently *a priori* power calculations showed 15 participants (in total) would be required to achieve an α error probability of 0.05 and power (1-β error) of 0.8. To ensure this sample size was achieved, accounting for drop-outs and those participants who may need to be excluded from analysis for not demonstrating having learnt the task, forty participants were targeted for recruitment.

6.3.2 Participants

Participants were recruited from the Participant Pool Scheme employed by the School of Psychology, University of Leeds. Such participants were chosen as being representative of the surgically-naive population, being drawn from students and staff at the University of Leeds. Participants with motor or cognitive impairment, impaired colour vision, previous surgical experience, or those who had participated in Experimentα, were excluded from the study. Participants were reimbursed £30 for their time.

6.3.3 Randomisation and Blinding

The order participants performed the preoperative simulation routines was determined by a stratified, permuted block randomisation sequence with a block size of 5 and an allocation ratio of 1:1. All assessment metrics were calculated by computer and consequently blind.

6.3.4 Ethical Approval

Experimentβ received ethical approval from the School of Psychology, The University of Leeds (ref: 15-0038, 04-Feb-2015 and 15-0199, 06-Aug-2015).

6.4 Results

Thirty-eight participants were recruited, with thirty-six undertaking all stages of the study (28 female : 8 male, average age 25 years [range 19 - 62], all (36) right-hand dominant).

6.4.1 Averaged Learning Results

All learning tasks (βLearn:K, βLearn:C and βLearn:C&K) demonstrated a consistent pattern of exponential improvement over successive trials: Figure: 6.4. As was found in Experimentα, the exponential function modelled the averaged results very well, as shown by high coefficients of determination (R^2): Table 6.1.

βLearn:C&K and βLearn:C appear to show very similar graphical features - they seem to demonstrate similar intercepts, rate of change and plateaus (represented by b, α and a in Equation 3.5, §3.3.1). Consequently, Akaike's Information Criterion (AIC) [139] was used to compare models. AIC is an estimation of the (relative) expectation of the Kullback-Leibler 'distance' (in actuality, probability distributions) between two

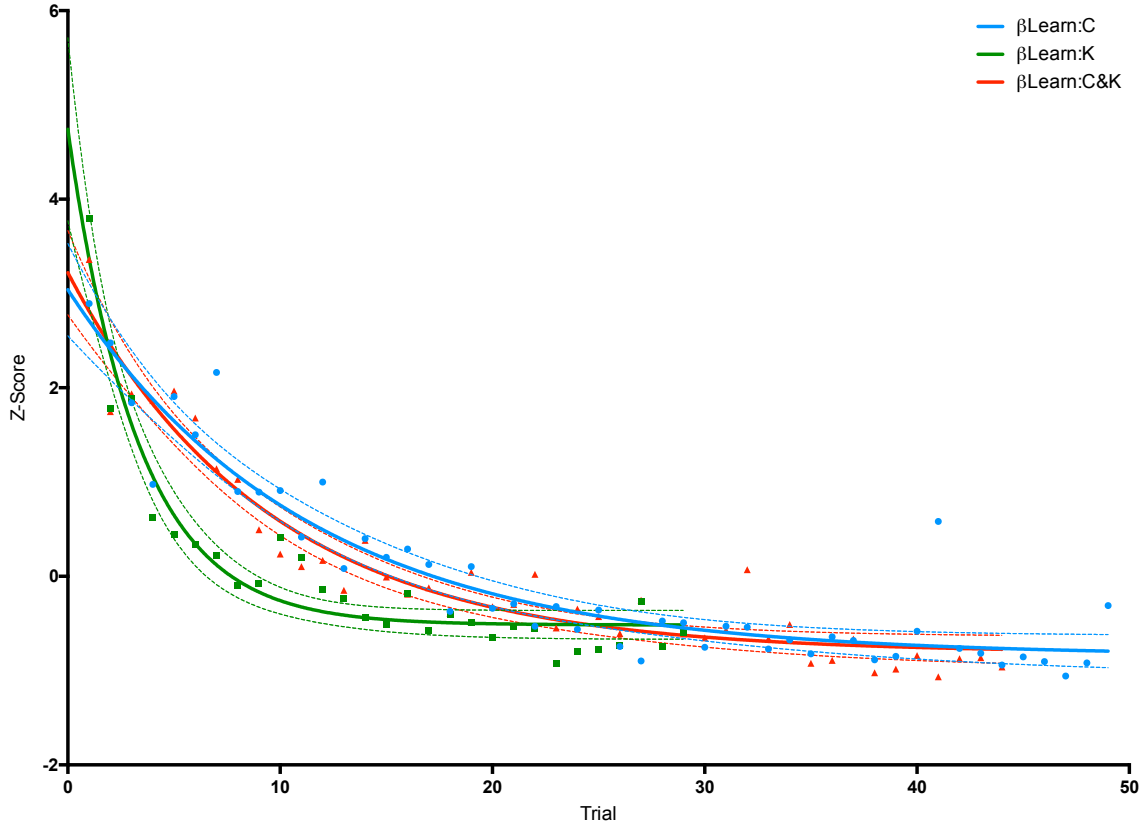


Fig. 6.4 Average Performance during the Experimentβ Learning Trials.
Learning Colour Information trials (βLearn:C) are represented in blue, Learning Kinematic Information trials (βLearn:K) are shown in green and Learning Colour and Kinematic Information trials (βLearn:C&K) in red. The dotted lines represent the 95% CI of the mean.
Note: Results are shown as dimensionless Z-scores to allow comparison between different trials.

Table 6.1 Coefficients of determination (R^2) for the averaged results of Experimentβ Learning Phase

Learning Trial	R^2
βLearn:C	0.8923
βLearn:K	0.9188
βLearn:C&K	0.9270

models, based on Fisher's maximised log-likelihood [140]. The Kullback-Leibler distance [141] in turn is the difference between two probability distributions P and Q, in which P represents the 'true' distribution of data, while Q represents the model of P. AIC is calculated using residual sums of squares from regression:

$$AIC = n{\cdot}ln\left(\frac{RSS}{n}\right) + 2K \tag{6.1}$$

Where n is the number of data points(observations), ln is the natural logarithm, RSS is the residual sums of squares and K is the number of parameters in the model.

AIC requires a bias-adjustment for small sample sizes, with a rule of thumb described by Burnham and Anderson [140] of if the ratio of n/K is less than 40, then use the bias-adjusted version of AIC - AIC_c, calculated thus:

$$AIC_c = n{\cdot}ln\left(\frac{RSS}{n}\right) + 2K + \frac{2K(K+1)}{n-K-1} \tag{6.2}$$

with the parameters being defined as above[1]. Calculating AIC_c for the individual models permits comparison between models and an estimation of how much better the best approximating model is compared with other model(s) [142]. The difference between values of AIC_c (Δ_i) can be used to calculate the Akaike weight, w_i, thus:

$$w_i = \frac{exp(-\frac{1}{2}\Delta_i)}{\sum_{r=1}^{R} exp(-\frac{1}{2}\Delta_r)} \tag{6.3}$$

The Akaike weight can be considered analogous with the probability that a given model is the best approximation [142]. Alternatively, the difference between models can be compared using the evidence ratio (ER):

$$ER = \frac{exp(-\frac{1}{2}\Delta_{best})}{exp(-\frac{1}{2}\Delta_i)} \tag{6.4}$$

The different models of learning are compared in Table 6.2.

Table 6.2 Comparison of individual and combined models of learning, βLearn:C and βLearn:C&K

Model	K	RSS	AIC_c	Δ_i	w_i	ER
Combined Curve	3	8.580	-213.2	0	0.875	
Individual Curves	4	4.672	-209.3	3.9	0.125	7.03

[1] As the size of a dataset (n) increases, the bias adjustment term becomes very small. Consequently, some authors argue that the small-sample adjustment (AIC_c) should always be used [142].

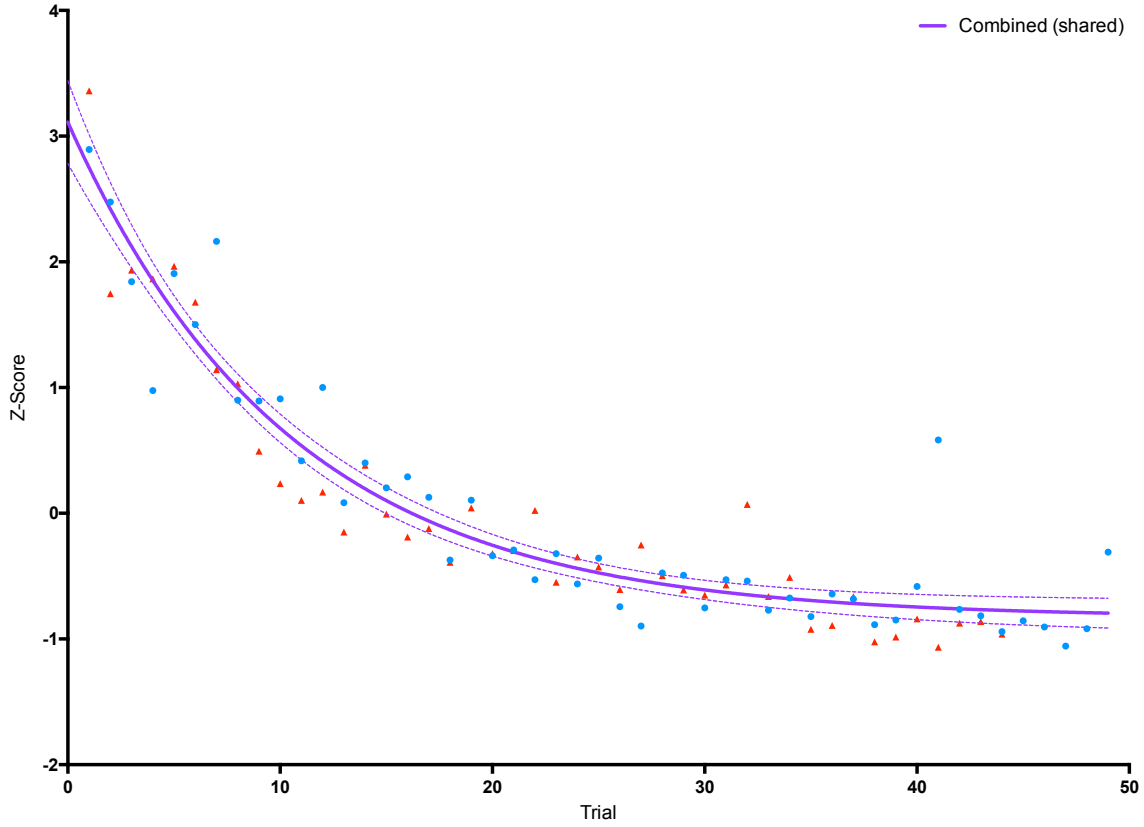


Fig. 6.5 Average Performance during βLearn:C and βLearn:C,&K with a combined (shared) learning curve. The dotted lines represent the 95% CI of the mean.

The results shown in Table 6.2 demonstrate that it is 7 times more likely that the results of βLearn:C and βLearn:C&K are explained by a single curve (87.5% likelihood) as opposed to two separate curves (12.5% likelihood). The combined curve is shown in Figure 6.5.

6.4.2 Individual Learning Results

Blocked individual learning results were compared to an exponential function of learning, as described in §3.3.1. The coefficients of determination, predicted plateaus of performance and the smallest (best) result achieved during the Learning Phase are reported in Appendix F; Tables F.1, F.2 and F.3.

The order participants undertook the learning trials did not have an effect on their performance at plateau: an independent samples t-test indicated no difference in performance at plateau for those participants who progressed through βLearn:K, βLearn:C and then βLearn:C&K learning tasks ($\bar{x} = -0.183$, $\sigma = 0.700$) when compared to those who performed βLearn:C, βLearn:K and finally βLearn:C&K ($\bar{x} = 0.1376$, $\sigma = 1.312$),

$t(23) = 0.883$, $p = 0.384$, $d = 0.305$. Levene's test [143] indicated unequal variances ($F = 4.80$, $p = 0.036$), so degrees of freedom were adjusted from 31 to 23.

6.5 Discussion

The results of βLearn indicate the majority of participants were able to learn at least one aspect of the laparoscopic sequence learning task, but that Experimentβ presented a greater challenge to learn than Experimentα. Mirroring the results of Experimentα, participants found learning kinematic information (βLearn:K, 26 participants successful) easier than the other two learning tasks. However, colour information (βLearn:C, 16 successful) appeared easier to learn than the combined colour and kinematic information (βLearn:C&K, 9 successful), in contrast to Experimentα.

Participants were more successful at learning colour information (βLearn:C) than the combined colour and kinematic information (βLearn:C&K) during βLearn, in contrast to Experimentα. One explanation for this findings may be because spatial information is not helpful in performing the Experimentβ task, as the spatial arrangement changes each time the task is completed. Consequently, while βLearn:C&K does provide participants with the required colour information necessary to complete ESOXβ, it does so in a way that is not as 'clean' as βLearn:C; βLearn:C&K requires participants to account for a kinematic transformation that is not required to complete βLearn:C. This appears to have acted as a confound for some participants trying to learn the colour-sequence information during βLearn:C&K. This hypothesis is further supported by the finding that those participants who were able to *successfully* learn βLearn:C&K showed *exactly* the same pattern of learning as those who were able to successfully learn βLearn:C (see Figure 6.5), suggesting these participants were learning the same information (i.e. colour-sequence information), in the same manner. For these participants, the kinematic transformation inherent in performing the task in a laparoscopic box trainer during βLearn:C&K (*cf.* βLearn:C which is performed with a pen-like stylus on a tablet-PC) appears to have had a negligible effect on their performance. Both tasks require the participants to plan and execute physical movements to complete, but βLearn:C&K disrupts the 1:1 ratio of movement of the upper limb to execution of the sequence found in βLearn:C. Some participants appear to have been able to account for this transformation. This finding suggests that the importance of the manner in which information is presented differs between discrete information modalities. For spatial information, incorporating a kinematic transformation is important when imparting information (see Chapter 5), whereas these learning trials indicate having to adjust to

a kinematic transformation makes no difference to colour information learning.

Chapter 7

Experimentβ: Assessment of Preoperative Simulation

After completing the learning phase, participants progressed to an assessment of preoperative simulation. Two forms of preoperative simulation were of benefit during Experimentα - αPS:C,K&S, which provided a specific preparation for participants, allowing them to practice exactly the same task before undertaking the assessed task and αPS:K&S which provided relevant information to the participants. This chapter follows a similar experimental design to α:PS.

7.1 Methods

Once they had progressed through the learning phase, participants were asked to perform five preoperative simulation exercises (including one control), one per day, in a randomly determined order, prior to performing an assessment task. This assessment task was very similar to the βLearn:C,K&S task described above, as the same spatial and colour information was maintained, but it did not give the participants any order information (there was no numbers in the squares): Figure 7.1. Each assessment was temporally separated by one day.

Participants were asked to perform one of each of the following preoperative simulation routines, in a randomly determined order, prior to performing the assessment task.

1. Kinematic Preoperative Simulation (βPS:K): Participants were asked to repeat Learning Task 1 (βLearn:K). This provided participants with kinematic (and spatial), but no colour information.

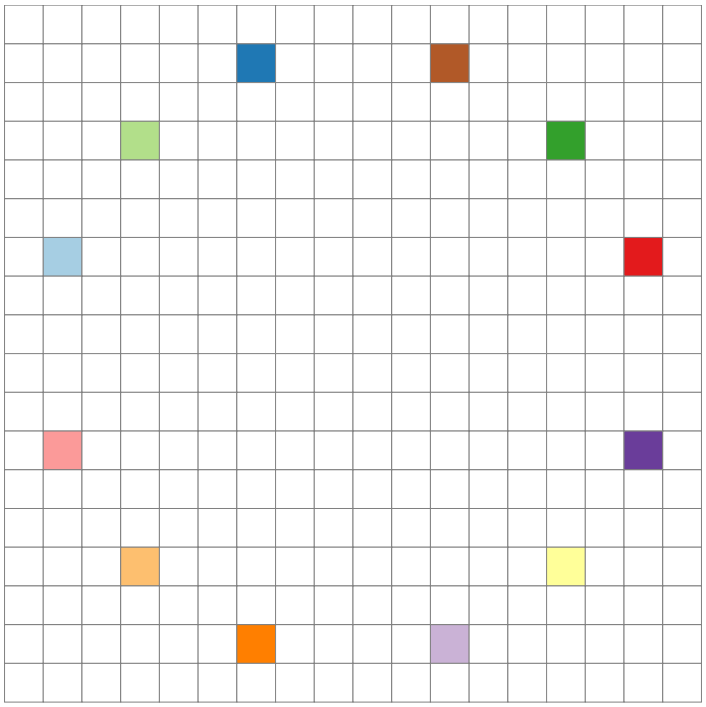


Fig. 7.1 Array for Assessment (without sequence information)

2. Colour Preoperative Simulation (βPS:C): Participants were asked to repeat Learning Task 2 (βLearn:C). This provided the participants with colour, but no kinematic information relating to the assessment task.

3. Observe Performance of Task (βPS:O): Participants were asked to view a screen-recording of ESOXβ being performed by an expert. This recording was performed by the author and was a 'perfect' performance of the task; no errors or deviations from the correct path were shown. This provided participants with colour information, but without direct physical involvement from the participant (no direct kinematic information). The recording was shown on the laparoscopic screen used to perform the ESOXβ task.

4. Colour and Kinematic Preoperative Simulation (βPS:C&K): Participants were asked to perform the sequence-learning task (βLearn:C&K), immediately prior to performing the assessment task (specific warm-up).

5. Control Group: Participants were asked to play Tetris®. This was chosen to occupy participants during the preoperative simulation period, without giving them any information about the upcoming assessment task and to provide a generic warm up of the visual-motor system. This preoperative simulation routine was performed on the laparoscopic stack using a keyboard.

Each preoperative simulation trial was limited to three minutes. This was controlled by computer script in all cases, aside from the control group which was timed by a supervising investigator. Participants could repeat the preoperative simulation routine as many times as was their preference during this time. This time restraint, chosen from pilot data, was employed to ensure parity between different participants and preoperative simulation routines. To further ensure this, each participant was given the same written instruction before each preoperative simulation and assessment, which were very similar to those used in ESOXα: Appendix E.

The above preoperative simulation routines were designed to offer the participants varying amounts of colour, spatial and kinematic information. Repeated performance of the assessment task (βLearn:C&K) was included as the theoretical 'ideal' preoperative simulation.

7.2 Results

Three participants (Participant No. 18, 22 and 35) met the exclusion criteria because they did not reach a predicted plateau of performance at any of the βLearn trials. Two participants (Participant No. 1 and 27) were excluded because their performance during assessment was more than 3 standard deviations (5.8 and 10.3, respectively) away from other participants' performance. One further participant (Participant No. 5) was also excluded from analysis, for reasons detailed in §7.2.1.

Analysis of performance following preoperative simulation was assessed using a repeated-measures ANOVA. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2 = 65.9$, $p < 0.001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.694$). A main effect of preoperative simulation on performance, $F(2.78, 80.5) = 3.01$, $p = 0.0384$, $\eta_p^2 = 0.0942$, was demonstrated. Post-hoc analyses using Fisher's LSD revealed this was driven by significantly better performance following the observe (O) preoperative simulation when compared to control ($p = 0.0459$) and better performance after the colour and kinematic (C&K) preoperative simulation when compared to control ($p = 0.0229$). These results are summarised in Figure 7.2.

Once again, the order participants performed the learning trials did not affect their performance at assessment: an ANCOVA [between-subjects factor: preoperative simulation (Control, Observe, K, C and C&K); covariate: learning order] revealed no interaction between preoperative simulation and learning order, $F(1, 28) = 0.0268$, $p = 0.871$, $\eta_p^2 = 0.001$.

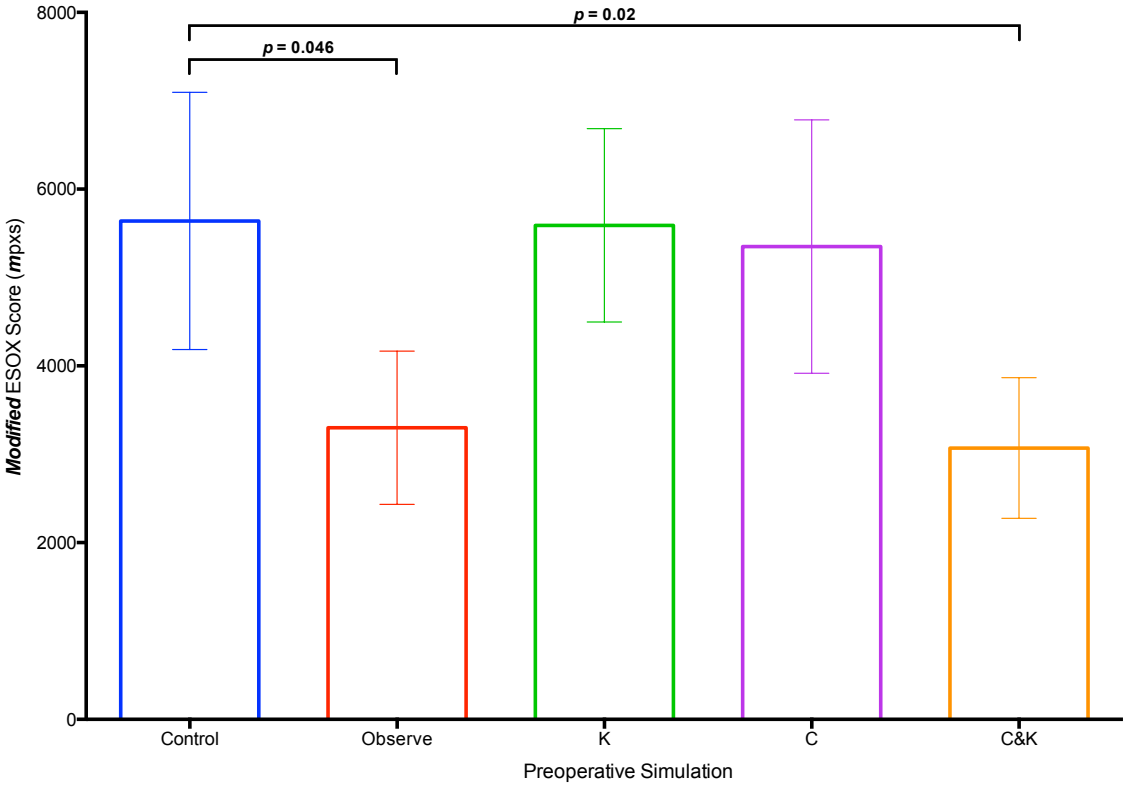
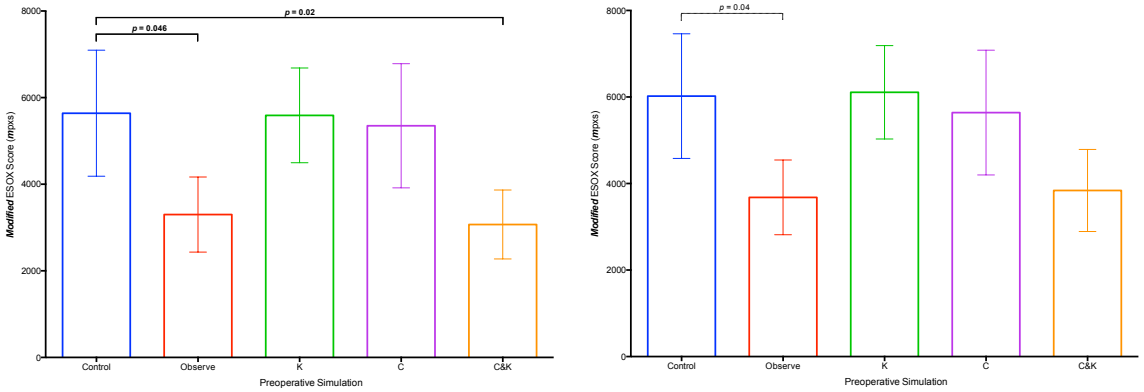


Fig. 7.2 Performance following Preoperative Simulation during Experimentβ. Significant within-subject effects are highlighted. The error bars represent the 95% CI of the mean. **Note**: Smaller scores represent better performance.



(a) Participant 5 excluded (as per Figure 7.2) **(b)** Participant 5 included

Fig. 7.3 Comparison of Performance following Preoperative Simulation during Experimentβ with and without Participant No. 5

7.2.1 Exclusion of Participant No. 5

An initial repeated measures ANOVA of the effects of preoperative simulation on subsequent assessed performance did not detect a main effect of preoperative simulation, but did closely approach statistical significance ($F(2.93, 87.9) = 2.59$, $p = 0.0594$, $\eta_p^2 = 0.0794$). Consequently, the individual results of participants during the assessment conditions were examined more closely. Overall, as shown in Figure 7.3b, the Observe and the Colour and Kinematic preoperative simulations appear to be the most beneficial for subsequent task performance. Looking specifically at these two conditions, the majority of participants exhibited similar patterns of results, with one exception; Participant No. 5 exhibited results following βPS:O and βPS:C&K that were 5.53 and 5.75 standard deviations away from all other participants, respectively. Interestingly, Participant No. 5 is a statistical outlier following all forms of preoperative simulation, with the exception of Colour, the only learning trial in which Participant No. 5 was able to demonstrate reaching a plateau of learning[1]: Figure 7.4. Further review of the learning trial results may demonstrate a cause for this finding; Participant No. 5 recorded the fewest number of learning trial repeats of any participant during Experimentβ. Consequently, while being able to satisfy the criteria for having learnt part of the Experimentβ task detailed in §3.3.1, Participant No. 5 may not have been truly able to satisfactorily learn the task, a conjecture supported by their *deterioration* of performance during βLearn:C&K.

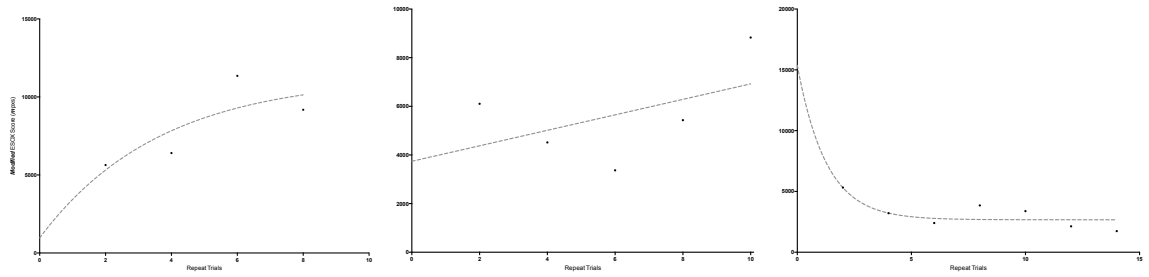
[1]Although Table F.2 appears to demonstrate that Participant No. 5 was able to reach a plateau of learning during βLearn:C&K, review of the individual learning graphs show that Participant No. 5 produced a *positive* exponential graph of repeated performance, i.e. their performance became *worse* with repetition (see Figure 7.4a).

Post hoc analysis of results is controversial, however, it seems justified in this instance, for the following reasons:

1. Review of Figures 7.3a and 7.3b reveals that the exclusion of Participant No. 5 did not alter that overall pattern of the effect of preoperative simulation. Both figures suggest that βPS:O and βPS:C&K improve performance relative to control, while βPS:C and βPS:K appear to have little effect.

2. The evidence that Participant No. 5 has effectively learnt the ESOXβ task is not as robust as that demonstrated by the majority of participants. Participant No. 5 was only able to demonstrate a plateau of repeat performances during one of the three learning trials (βLearn:C). During both βLearn:K and βLearn:C&K, Participant No. 5's performance *worsened* with repetition, counter to the majority of participants and expectations: Figures 7.4a and 7.4b. Participant No. 5 was able to meet the criteria for learning during βLearn:C, however, as described in §3.3.1. Although an exponential decrease in performance is expected (and has robust theoretical and literature support) if learning is occurring with repeated trials, the threshold goodness-of-fit measure ($R^2 \geqslant 0.5$) was chosen arbitrarily. This was done because no agreed standard of threshold measurement exists (and is likely to be dependent on the task being assessed). While strict adherence to an exponential model ($R^2 \gtrapprox 0.8$) is very likely to represent learning and the converse of no exponential improvement ($R^2 \lessapprox 0.1$) is likely to demonstrate a lack of learning, an intermediate R^2 is much more difficult to interpret, particularly with the paucity of data provided by Participant No. 5 (given their limited repeat performances). In addition, review of a graph of repeat performances during βLearn:C with a calculated exponential curve of performance improvement (Figure 7.4c) seems to show that Participant No. 5 does not demonstrate an initial significant performance improvement. These factors in combination - a worsening of performance during two of the learning trials, the lowest number of repeat trials of any participant during βLearn and a questionable goodness-of-fit measure - undermine the initial finding that Participant No. 5 successfully learnt βLearn:C. As shown previously [104], learning is a significant confound of preoperative simulation.

3. Participant No. 5's results at assessment of preoperative simulation were significantly divergent from all other participants. Preoperative simulation has been assessed throughout using a repeated-measures ANOVA, in order to examine between-subjects effects and allow for differing levels of performance by individ-



(a) βLearn:C&K **(b)** βLearn:K **(c)** βLearn:C

Fig. 7.4 Participant No. 5's performance during the βLearn trials. The dotted lines are an attempt to fit an exponential function (of learning) to successive trials.

ual participants. Consequently, a significant variation in the level of performance achieved by individual participants can be accounted for. However, Participant No. 5 produced results that are so far removed from those of all other participants (more than five standard deviations away from the population following βPS:O and βPS:C&K) it suggests entirely different processes are occurring when Participant No. 5 undertakes the preoperative simulation routines when compared to all other participants, which may be as a result of incomplete learning, as described above.

These factors; significantly divergent results during assessment of preoperative simulation and apparent lack of having learnt the ESOXβ task satisfactorily, demonstrate compelling evidence for the necessity of excluding Participant No. 5.

7.3 Discussion

The results of Experimentβ reveal preoperative simulation to have a significant effect on subsequent performance. This effect is driven by an improvement in performance following an observation of expert performance and a similar improvement following a colour and kinematic preoperative simulation. It seems intuitive that repeat performance of a task (as provided by βPS:C&K) would result in subsequent performance improvement and is discussed in more detail below in §7.4. That βPS:O was beneficial whereas βPS:C has no effect is an interesting finding. Experimentβ was designed to give predominance to colour information, with spatial information being randomised each time Experimentβ was performed. It was consequently expected that colour information would be a beneficial form of preoperative simulation (particularly compared with kinematic information alone, which was shown to be equivalent to the control condition during the assessment of PS). However, it appears that the way in which

colour information is presented is critical. βPS:C and βPS:O contain the same colour information, i.e. that Square 1 is peach-coloured, Square 2 is pale green, Square 3 is red, etc. However, βPS:C requires participants to actively demonstrate their understanding of the colour order by moving a pen-like stylus to each coloured square, in order, on a tablet computer. βPS:O only required participants to watch a recording of expert performance. One pre-test hypothesis was that βPS:C would be a more beneficial form of preoperative simulation as it necessitates the active involvement of participants, whereas participants could choose to ignore or not actively engage with βPS:O. The findings of Experimentβ do not support this prediction. This may be because βPS:C requires the utilisation of multiple neural processes (both higher-order cognitive processes involved in the interpretation of colour and sequence information but also movement planning and execution processes necessary for moving the stylus to the right coloured square on the tablet computer), whereas βPS:O provides the requisite colour sequence information in a 'cleaner', less cognitively demanding manner, i.e. without the need to consider subsequent movement.

This is similar to the conjecture, described in §6.5, for the reason that βLearn:C seems to have been easier to learn than βLearn:C&K and again mirrors the findings of Experimentα.

7.4 Experimental Comparisons

Both Experimentα and Experimentβ demonstrate preoperative simulation has an effect on subsequent performance. In both experiments, two preoperative simulation routines were shown to be effective; the kinematic and spatial PS and the colour, kinematic and spatial PS in Experimentα and the observe PS and the colour and kinematic PS during Experimentβ.

That repeated performance of a task would result in an improvement in performance of that task seems intuitive. αPS:C,K&S allowed participants to practice *exactly* the same task as they were subsequently assessed on after three minutes. βPS:C&K allowed participants to practice *almost* exactly the same task as they were subsequently assessed on, with the difference of a random spatial arrangement, within the constant circular design, each time the task was performed. These findings support the notion that performing the same laparoscopic operation twice could lead to improved performance during the second case (see Chapter 10 for an exploration of this effect in clinical practice). This is the same finding as the majority of studies systematically reviewed in Chapter 2. However, the practical application of this finding is questionable as

two of the same operations can vary significantly in the real world. Exactly the same operation can be recreated within a simulator, but the ability to generate a simulation of a procedure for a specific patient prior to performing that operation is not currently technically feasible. With improvements in simulation technology and non-invasive imaging modalities, it is unlikely that such a capability will remain theoretical for very long. Alternatively, it may be that structuring operating lists to involve the repetition of the same procedure produces a tangible benefit (though this would not help the first patient on an operating list). However, it is the other finding of Experimentα and Experimentβ that is arguably more interesting - that performing other simplified, relevant procedures/tasks prior to a simulated operation can improve a participant's performance during that operation.

Experimentα was designed to give equal weighting to both spatial and kinematic information, with participants being able to successfully complete the task using either colour information or spatial information or a combination of both modalities. The results of αLearn (comprising of all the learning trials within Experimentα; αLearn:K&S, αLearn:C and αLearn:C,K&S) demonstrate that participants had a preference for kinematic and spatial information, with the subsequent finding of αPS that a kinematic and spatial preoperative simulation was beneficial. The reason for this preference is likely to reflect the preference of human cognitive processing for the interpretation and storage of spatial information compared with more abstract colour processing.

Experimentβ was designed to necessitate the use of colour information to complete the laparoscopic sequence learning task. Successfully learning this information was more difficult than learning the spatial information preferred in Experimentα, but once learnt, observing expert performance of the task (which included the correct colour sequence information) significantly improved subsequent performance. The efficacy of βPS:O is likely a reflection of the fact that βPS:O provided participants with a reminder of the the requisite colour information in the 'simplest' way (in particular when compared to βPS:C).

These findings would also suggest that active engagement of the motor system is important in facilitating performance that relies on spatial information - αPS:K&S was most effective at preparing participants to perform the Experimentα task. However, such active engagement of motor systems is not required for other components of visual information - namely colour. This may reflect the differences in spatial and colour neural processing [144].

In both Experimentα and Experimentβ the preoperative simulations that provided participants with the necessary information to complete the assessed task in the most simple manner were found to be beneficial. In both experiments, the nature of the

advantageous preoperative simulation is dependent on the nature of the task being performed and is likely to reflect the cognitive strategies employed in performing that task (which in turn are likely to reflect how that task was learnt).

Chapter 8

Experimentγ: Learning-Phase

Experimentα and Experimentβ have demonstrated that preoperative simulation can have a significant impact on subsequent performance. It seems that simplified, relevant procedures/tasks are of benefit as a preoperative simulation, which holds great promise for the use of preoperative simulation in clinical practice. The Experimentα and Experimentβ tasks engendered a distinct preference for one type of information in participants; spatial information during Experimentα and colour during Experimentβ, and this was subsequently reflected in the preoperative simulation routines that were beneficial. Experimentα could be completed by utilising spatial information, colour information or a combination of both, with participants demonstrating a preference for spatial information during both the learning and assessment of preoperative simulation phases. Experimentβ could only be completed by utilising colour information, forcing participants to use this information during the learning phase and subsequently it formed an effective preoperative simulation.

Relating these abstract experiments back to MIS, Experimentα demonstrates that if a surgeon could complete a laparoscopic operation by utilising spatial information *alone*, a spatial preoperative simulation would be beneficial, i.e. if an operation could be completed successfully using technical ability alone, a technical preoperative simulation would be helpful. However, under such an experimental design, more abstract knowledge (represented by colour) is of no benefit.

Experimentβ explored the effect of a knowledge-based preoperative simulation under conditions in which knowledge (colour) was paramount. In this scenario, a technical (spatial) preoperative simulation was of no benefit, but a revision of previously learnt information (colour) was helpful as a preoperative simulation.

Experimentγ, the third randomised, cross-over study utilising the ESOX program, has been designed to create a more balanced distribution of importance between colour

and spatial information, to try and better reflect clinical practice; Anatomical variations are common[1] [145, 146] and the attending surgeon needs to have an understanding of both 'normal' anatomy and the common deviations. To return to the crude analogous aspects of the experimental descriptors discussed in §6, spatial information relates to the physical actions of operating while colour information relates to the more abstract, academic knowledge required to perform surgery. Both of these aspects are required; an operation cannot be completed by only relying on either in isolation. The previous two experiments have examined each of these aspects in turn, but not when both are relied upon by the participants.

In order to achieve a situation in which both colour and spatial information are utilised, the reliability of spatial information must be degraded, as described below. Here the aim was to make spatial information useful, but not to the exclusion of colour information.

8.1 Experimental Task γ

Experimentγ is another randomised, controlled cross-over trial that utilises a sequence-learning task developed using the ESOX program. It exhibits many similar characteristics to the two previous iterations (Experimentα and Experimentβ), but is designed to deliver spatial and colour information in a manner that ensures both are useful. Utilising both colour and spatial as independent, though related, sources of information, results in a preference for spatial information; see Experimentα, Chapters 4 and 5. It is very difficult in such a design to *increase* the importance of colour information and make it more likely that participants would use colour information in order to complete the task. However, as shown in Experimentβ (Chapters 6 and 7), participants will use colour information to complete the laparoscopic sequence-learning task if they are unable to rely on spatial information. This would suggest there is a point between the two previous distributions of information at which both forms of information would be of equivalent benefit to participants. In order to achieve such a balance, the integrity of spatial information has to be degraded.

In Experimentγ, participants were shown a sequence of eight numbered, coloured squares distributed on a 14 by 28 array and were asked to move a laparoscopic stylus to each square, in order. Though participants are initially shown the correct sequence order, once a trial is started by the participant, the numbers are removed, necessitating the participants to remember the sequence order of coloured squares. To aid learning

[1] For example, a recent review of over one thousand magnetic resonance cholangiopancreatography images revealed over forty percent demonstrated anatomical variations of the biliary tree [145].

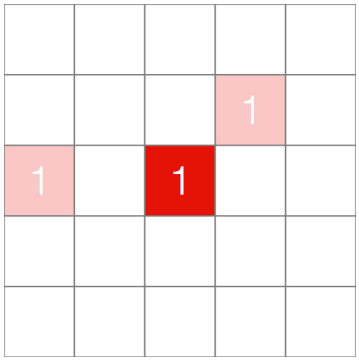


Fig. 8.1 Example of the fuzzy location parameter employed in Experimentγ. The original (primary starting) location is shown in dark red, with two possible subsequent locations shown in lighter red. The square has been programmed to move ± 2 squares.

and provide feedback, once the correct square is reached, it 'disappears', informing participants that they have reached the correct square. Colour information was maintained throughout each repetition, but the spatial location of squares could change randomly each time a trial was conducted. Each individual square could be programmed to 'walk' a certain distance from its primary starting location (PSL) at the beginning of each trial. This 'walk' is actually a fuzzy location parameter [147], allowing squares to move a predetermined distance in either the x or y (or both) direction(s), before being made visible to participants: Figure 8.1.

The number of coloured squares in the sequence utilised was reduced from twelve in Experimentβ to eight, to try and improve the learning phase of the experiment. This was in light of the previous experiments, during which some participants failed to learn the complete sequence and did not reach a plateau of performance, and in view of the added complexity of an uncertain spatial arrangement. Additionally, the colours used for each square was again modified to try and make them more categorical i.e. a distinct colour that could be named (such as red, blue) as opposed to varying shades of colours. The selection of colours were made using the cartographical tool ColorBrewer [138].

8.2 Methods

As during Experimentα and Experimentβ, participants progressed through three stages of training during week one of Experimentγ:

1. Learning Colour Information (γLearn:C): in this task, colour information was provided and maintained through each repetition (i.e. Square 1 was always red), but the position of each square with a circular arrangement was randomly deter-

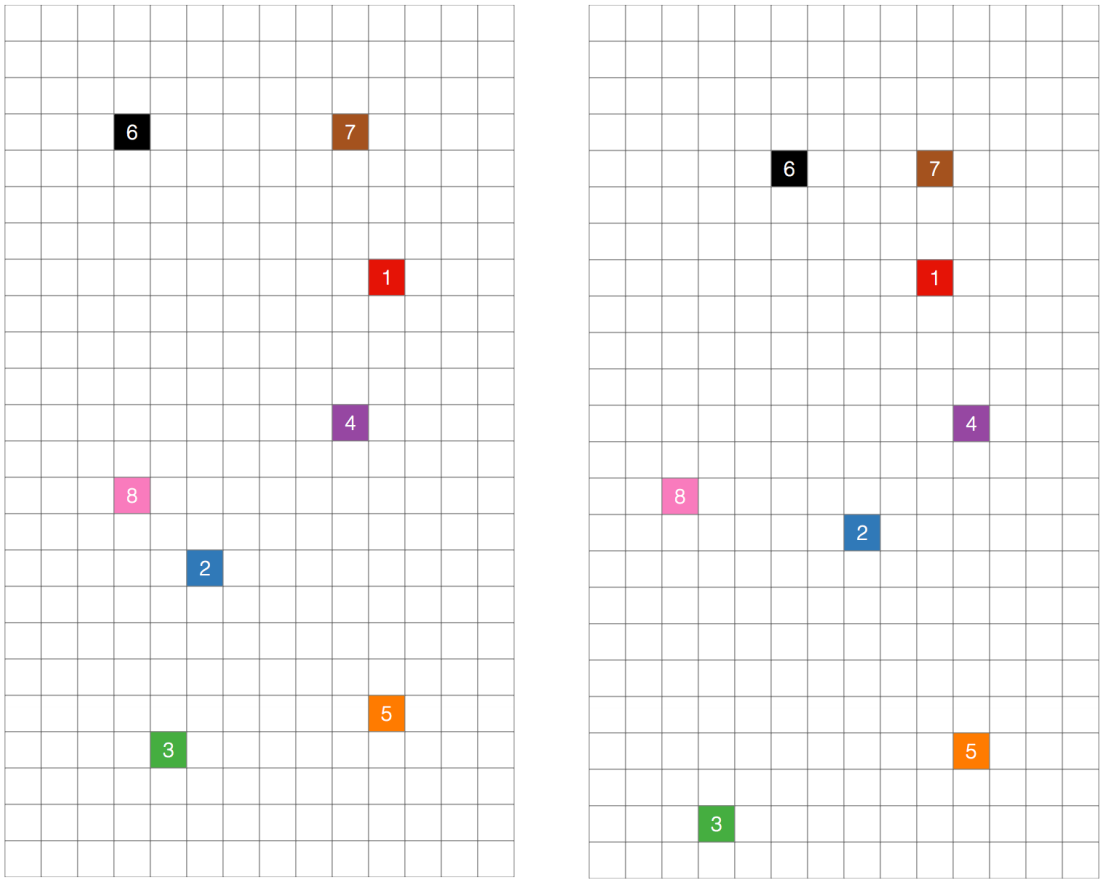
mined by applying a Fisher-Yates shuffle [97] before each array was built (as per αLearn:C and βLearn:C; see Figures 4.2 and 6.3). This task was performed on a tablet computer using a normal pen-like stylus. It was employed to promote learning of the colour sequence, independent of spatial location.

2. Learning Colour, Kinematic and Spatial Information (γLearn:C,K&S): during this task, coloured squares were distributed in a set spatial pattern across the experimental array: Figure 8.2a. The spatial location of squares did not change with each repetition of the task and the colour information was maintained from γLearn:C. As in previous experiments, this task was performed on a tablet computer within a laparoscopic box trainer using a laparoscopic stylus: Figure 3.2. This task was utilised to promote the learning of spatial information (see Experimentα).

3. Learning Colour, Kinematic and *Fuzzy* Spatial Information (γLearn:C,K&*F*S): in this task, coloured squares were distributed in a fuzzy set [147] pattern across the experimental array: Figure 8.2b. In this task, the location of each individual square could vary from the set position found in γLearn:C,K&S by ± 3 squares, in either the horizontal, vertical or both plains. To reinforce the 'correct' distribution of squares (PSL), participants alternated between fuzzy spatial and regular, 'non-fuzzy' spatial information. As demonstrated in Figures 8.2a and 8.2b, the general location of a square could not change dramatically - square 3 will remain in the bottom left hand corner of the array, but the precise location of squares will change with each repetition. This task was performed on a tablet computer within a laparoscopic box trainer using a laparoscopic stylus. It was designed to promote the learning of *both* colour and spatial information.

As with Experimentα and Experimentβ, participants were randomised in groups to receive either Task 1 then Task 2, or Task 2 and then Task 1. Subsequently, all participants completed Task 3 (γLearn:C,K&*F*S), to control for any effects of learning order. To further try and aid learning, after each trial (for all tasks), participants were given feedback about their performance from the ESOX program in the form of time taken and distance travelled in completing the task. Such feedback has been demonstrated to improve student learning of technical surgical skills [115–118]. This information was displayed on the screen after each task was completed.

All participants were trained to perform Task One, Two and Three in two separate training sessions. These training sessions were temporally limited to a maximum of one hour. As participants had to learn three tasks, they were encouraged to spend



(a) γLearn:C,K&S **(b)** γLearn:C,K&FS

Fig. 8.2 Arrays for γLearn

approximately 20 minutes on each. However, this was not fixed. If participants felt they had learnt the tasks prior to one hour, the session could be terminated early by the participant. Sessions were performed on separate days to enable off-line processing and consolidations of the tasks [57–60].

8.3 Analysis

Analysis of Experimentγ mirrored that of Experimentα and Experimentβ, as detailed in §4.3 and §6.3. Learning was assessed by comparing participants' performance to an expected negative exponential function and the effects of preoperative simulation were analysed using a repeated measures ANOVA.

8.3.1 Ethical Approval

The Experimentγ study received ethical approval from the School of Psychology, The University of Leeds (ref: 16-0151, 24-May-2016).

8.4 Results

Forty-three participants were recruited, with thirty-four undertaking all stages of the study (23 female : 11 male, average age 28 years [range 18 - 64], 26 right-hand : 8 left-hand dominant). The results of preoperative simulation are discussed in further detail in §9.2. Three participants dropped out of the study (Participant No. 4, 8 and 9) and three were unable to demonstrate having learnt any of the γLearn trials (Participant No. 16, 38 and 40).

8.4.1 Averaged Learning Results

All learning tasks (γLearn:C, γLearn:C,K&S and γLearn:C,K&FS) demonstrated a consistent pattern of exponential improvement over successive trials: Figure 8.3. During the γLearn:C,K&FS trial participants alternated between fuzzy and 'normal' spatial information, and consequently, these results are presented separately (γLearn:F_0 without fuzzy information, γLearn:F_1 with fuzzy spatial information; Figure 8.3 insert) and in combination (γLearn:C,K&FS). As was found during the previous two experiments, the exponential function modelled the results well: Table 8.1.

Mirroring the results of Experimentβ, γLearn:C and γLearn:C,K&S seem to demonstrate very similar features - apparently comparable intercepts, rates of change and plateaus, as do γLearn:F_0 and γLearn:F_1. Comparing these seemingly equivalent curves using AIC_c [139], reveals that γLearn:F_0 and γLearn:F_1 are best explained by a single curve while γLearn:C and γLearn:C,K&S are more accurately explained using individual curves: Tables 8.2 and 8.3. These results show that it is more than sixteen times more probable that the results of γLearn:F_0 and γLearn:F_1 are explained by a single curve (94.3% likelihood), when compared to two separate curves (5.7% likelihood). Conversely, γLearn:C and γLearn:C,K&S are three times more likely to be explained by individual curves.

8.4.2 Individual Learning Results

Blocked individual learning results were compared to an exponential function of learning, as described in §3.3.1. The coefficients of determination, predicted plateaus of

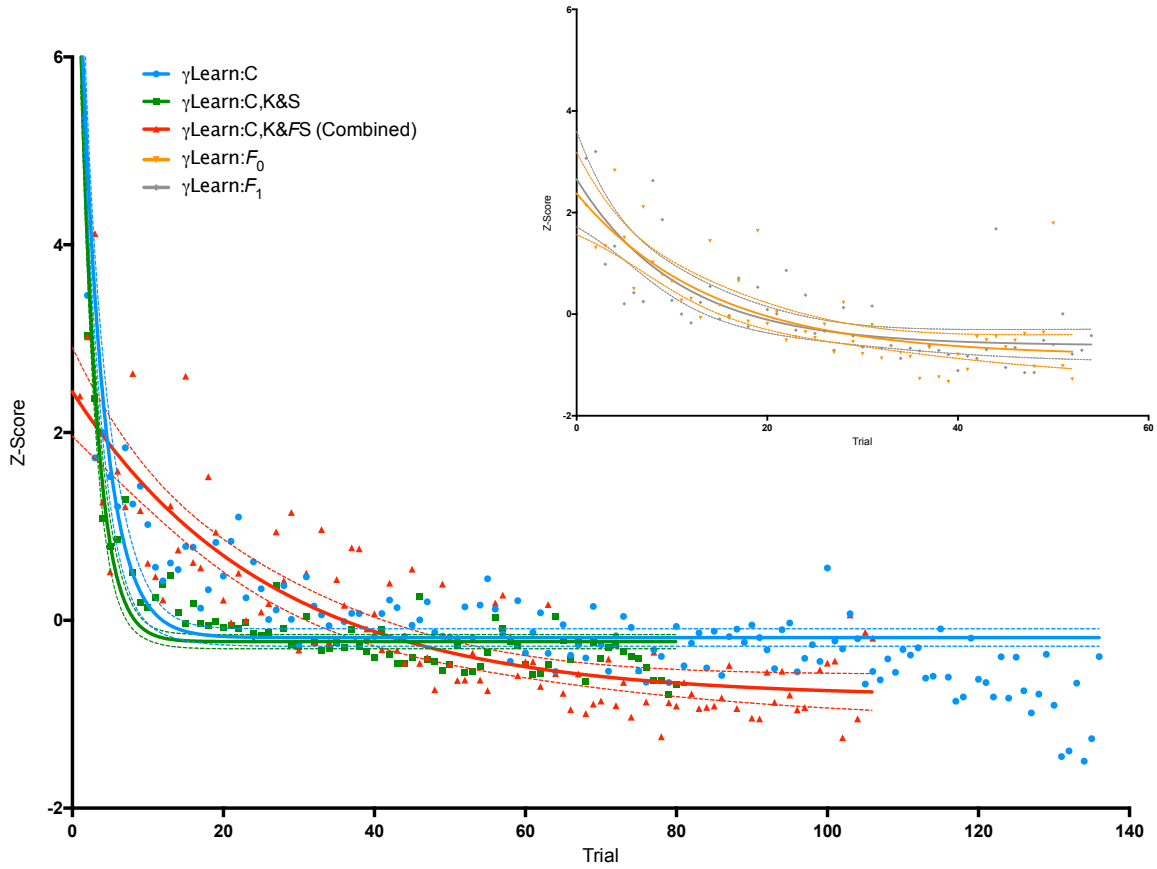


Fig. 8.3 Average Performance during the Experimentγ Learning Trials.
Learning Colour Information trials (γLearn:C) are represented in blue, Learning Kinematic Information trials (γLearn:C,K&S) are shown in green and Learning Colour and Kinematic Information trials (γLearn:C,K&FS) in red.
γLearn:C,K&FS is further subdivided into trials in which fuzzy spatial location (F_1, grey) and without fuzzy locations (F_0, orange), insert.
The dotted lines represent the 95% CI of the mean.
Note: Results are shown as dimensionless Z-scores to allow comparison between different trials.

Table 8.1 Coefficients of determination (R^2) for the averaged results of Experimentβ Learning Phase

Learning Trial	R^2
γLearn:C	0.7252
γLearn:C,K&S	0.8997
γLearn:C,K&FS	0.7117
γLearn:F_0	0.7530
γLearn:F_1	0.6843

Table 8.2 Comparison of individual and combined models of learning, γLearn:F_0 and γLearn:F_1

Model	K	RSS	AIC_c	Δ_i	w_i	ER
Combined Curve	3	27.91	-128.4	0	0.943	
Individual Curves	4	13.8	-122.8	5.6	0.0567	16.63

Table 8.3 Comparison of individual and combined models of learning, γLearn:C and γLearn:C,K&S

Model	K	RSS	AIC_c	Δ_i	w_i	ER
Combined Curve	3	46.8	-322.0	0	0.250	2.99
Individual Curves	4	22.5	-319.8	-2.188	0.740	

performance and the smallest (best) result achieved during the Learning Phase are reported in Appendix G; Tables G.1, G.2 and G.3.

The order participants undertook the learning trials did not have an effect on their performance at plateau: an independent samples t-test indicated no difference in performance at plateau for those participants who progressed through γLearn:C, γLearn:C,K&S and then γLearn:C,K&FS ($\bar{x} = 10.2$, $\sigma = 2.07$) when compared to those who performed γLearn:C,K&S, γLearn:C and finally γLearn:C,K&FS ($\bar{x} = 9.51$, $\sigma = 2.74$), $t(41) = 1.01$, $p = 0.384$, $d = 0.318$.

8.5 Discussion

The results of γLearn are similar to those of αLearn and βLearn; the majority of participants were able to learn at least one aspect of the simulated laparoscopic task, but the task remained a significant challenge to participants, with only two participants demonstrating they reached a clear plateau of performance in all of the learning trials. Once again, when participants could rely on spatial information to complete the task (γLearn:C,K&S), they performed better. When spatial information was uninformative (γLearn:C) or unreliable (γLearn:C,K&FS), participants found the task much harder to learn, as evidenced by the fewer number who reached a plateau of learning.

The two aspects of γLearn:C,K&FS, γLearn:F_0 and γLearn:F_1, are best explained by a single learning curve. As participants alternated between γLearn:F_0 and γLearn:F_1 during γLearn:C,K&FS it may have been the case that participants switched 'modes' of performance between each repeat and interacted differently with the primary layout when compared to the fuzzy layout. However, it appears they *are* the product of a single learning curve, possibly suggesting that participants used the same strategy to

complete both iterations of the task (with and without fuzzy spatial locations). This also acts as a demonstration of the usefulness of AIC_c in investigating comparable learning curves.

γLearn:C and γLearn:C,K&S are best described by individual curves, suggesting these tasks are learnt in different ways, utilising different information. This could suggest that while participants had to use colour information to complete γLearn:C, different information was utilised to learn γLearn:C,K&S. It could be argued that the different methods of completing the two tasks resulted in different learning curves (a laparoscopic stylus and box trainer during γLearn:C,K&S and a pen-like stylus on a tablet during γLearn:C), but such an argument is contradicted by the findings of βLearn; βLearn:C and βLearn:C&K demonstrated a shared curve of learning, suggesting the same information was being learnt, despite different methods of executing the tasks. Therefore, it seems reasonable to assume that colour information was utilised exclusively during γLearn:C (other information was not helpful), but spatial (or a combination of spatial and colour) was used during γLearn:C,K&S, echoing the findings from αLearn - participants will preferentially utilise spatial information, if able to do so.

Chapter 9

Experimentγ: Assessment of Preoperative Simulation

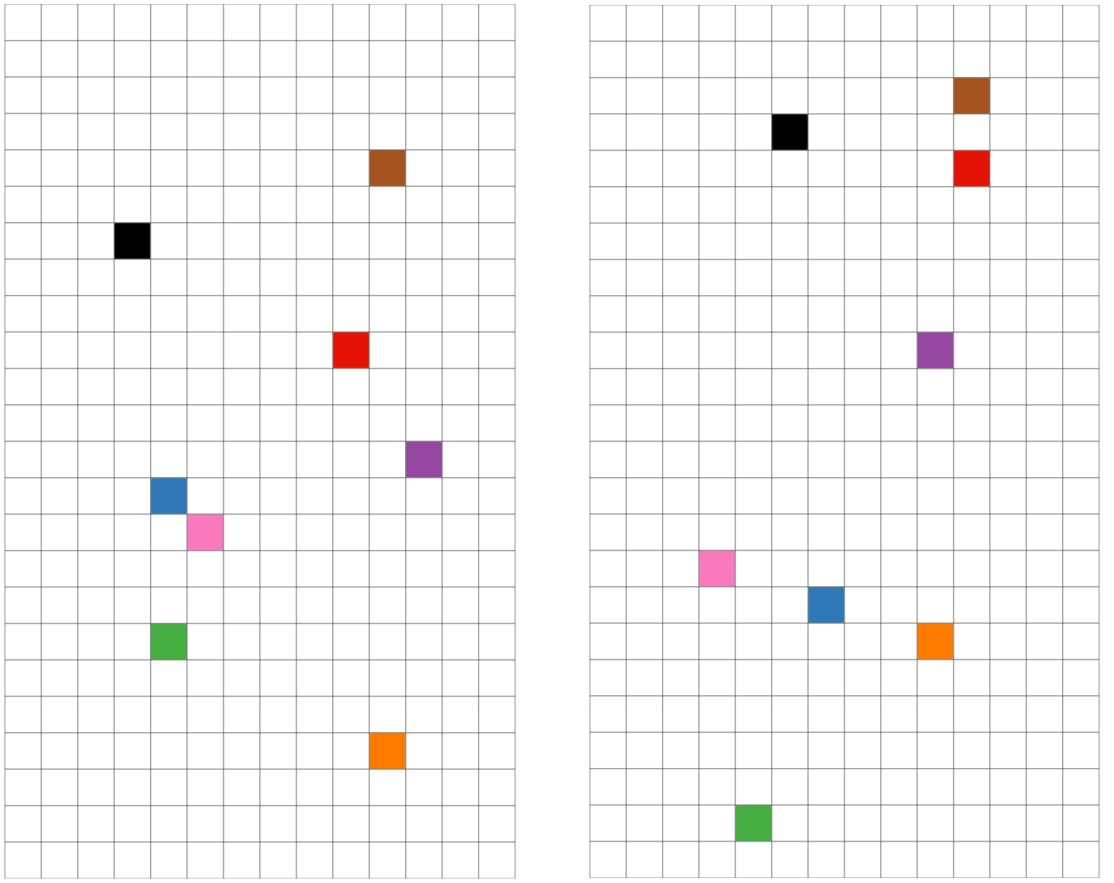
Following the learning phase of Experimentγ, participants performed an assessed 'procedure' after experiencing a variety of preoperative simulations. Experimentα and Experimentβ have shown that repeating the same task prior to assessed performance is beneficial but also that providing participants with information which is relevant to completing the task is also of significant benefit. Experimentγ follows a similar design to the preceding two experiments. Consequently, the beneficial forms of PS from Experimentα and β were assessed, along with an investigation into the effect of task variation on preoperative simulation.

9.1 Methods

Having proceeded through the learning phase, all participants were asked to perform five assessments with a different preceding preoperative simulation exercise (including one control), one per day, in a randomly determined order. This assessment phase of the experiment took place in the week following the learning phase. The assessment task was very similar to γLearn:C,K&FS, but it did not give the participants any order information (there were no numbers in the squares): Figure 9.1. Each assessment was temporally separated by one day.

Participants were asked to perform one of each of the following preoperative simulation routines, in a randomly determined order, prior to performing the assessment task.

1. Colour, Kinematic and Spatial Preoperative Simulation (γPS:C,K&S): Participants were asked to perform Learning Task 2 (γLearn:C,K&S). This provided the



(a) Array for Assessment (Config. 1) **(b)** Array for Assessment (Config. 2)

Fig. 9.1 Array for Assessment, demonstrating different configurations with fuzzy spatial location

participants with colour and spatial information, as well as information about the kinematic transformation. The task remained unchanged with each repetition. This task was undertaken on a tablet computer within a laparoscopic box trainer using a laparoscopic stylus (Figure 3.2).

2. Observe Performance of Task (γPS:O): Participants were asked to view a screen-recording of Experimentγ being performed by an expert. This recording was performed by the author (who is highly proficient at the task) and was an essentially ideal performance of the γPS:C,K&FS task; no errors or deviations from the correct path were shown. This provided participants with colour and spatial information, but without direct physical involvement from the participant (no direct kinematic information). The recording was shown on the laparoscopic screen used to perform the experimental task.

3. Colour, Kinematic and constant *Fuzzy* Spatial Preoperative Simulation (γPS:-

C,K&*F*S): Participants were asked to perform the third learning task (γLearn:-C,K&*F*S), immediately prior to performing the assessment task. The conditions of fuzzy spatial arrangement (i.e. the distance the squares could 'walk') did not change with repetition of the task, squares continued to be able to move ± 3 squares in the x and/or y dimension from their stating location. Note: the starting location (PSL) remains as per γLearn:C,K&S / γPS:C,K&S.

4. Colour, Kinematic and *Random Fuzzy* Spatial Preoperative Simulation (γPS:-C,K&*RF*S): this preoperative simulation is very similar to the above γPS:C,K&*F*S, but the degree of fuzzy spatial distribution varied with each repetition - squares could be distributed from ± 1 to ± 5 from their starting locations (PSL). This task was performed on a tablet computer within a laparoscopic box trainer using a laparoscopic stylus.

5. Control Group: Participants were asked to play Tetris®. This was chosen to occupy participants during the preoperative simulation period, without giving them any information about the upcoming assessment task. This preoperative simulation routine was performed using the same monitor as the laparoscopic stack utilised in previous trials, but was completed using a keyboard.

Each preoperative simulation trial was limited to three minutes. This was controlled by computer script in all cases, aside from the control group which was timed by a supervising investigator. Participants could repeat the preoperative simulation routine as many times as was their preference during this time. This time constraint was based on pilot data and was employed to ensure parity between different preoperative simulation routines. To further ensure equivalent conditions, each participant was given the same written instructions before each preoperative simulation and assessment, which were very similar to those used in Experimentα: Appendix E.

The first three forms of preoperative simulation were chosen as they were shown to be effective during Experimentα and Experimentβ; repetition of the task was beneficial during the previous two experiments, as was the PS that provided relevant information to the participants (αPS:K&S and βPS:O). The final preoperative simulation (γPS:C,K&*RF*S) was developed to investigate the effect of task variation in preparing a surgeon to operate. Varying a task during learning has been shown to improve performance at subsequent novel (though related) tasks [104, 148–151]. When presented with a novel task after experiencing random variation of similar tasks, participants demonstrate key components of structural learning: a facilitation of learning tasks with the same structure, a significant reduction in the interference that typically

occurs when switching between tasks that have alternative control strategies, and a preference to explore a novel task along the previously learnt structure [149]. Consequently, to investigate if such an effect could also be demonstrated in preoperative simulation, γPS:C,K&RFS was intended to introduce further variation into the laparoscopic sequence learning task. Some variation exists within the assessment task due to the fuzzy location parameter (squares could vary in location by ± 3). However, the amount of variation is fixed and squares will remain in the same basic location as their PSL (i.e. Square 3, Green will remain in the bottom left corner of the array during each repetition). However, γPS:C,K&RFS breaks both of these conventions by varying the amount that squares can move (from one - five squares in the x or y direction), which can result in squares entirely changing their location (Square 3 could be in the top left corner). γPS:C,K&RFS aims to determine if increased exploration can improve preparation in a similar manner that it aids learning.

In relation to clinical practice, γPS:C,K&RFS aims to explore the effect of variation in preoperative simulation. This is crucial as the majority of current preoperative simulation routines utilise unchanging simulations (see Chapter 2). This have been criticised [152] as not being representative of the surgical population, where anatomical variations are common [145, 146].

9.2 Results

Three participants (Participant No. 4, 8 and 9) did not complete any assessment of PS as they chose not to finish the study. Three participants (Participant No. 16, 38 and 40) were unable to demonstrate having reached a plateau of performance during any of the γLearn trials and were consequently excluded. Six participants (Participant No. 5, 14, 21, 27, 34 and 42) did not complete all preoperative simulations and consequently had to be excluded from initial analysis, but four did complete four of the five assessments of PS and so were included in planned comparisons. Three participants (Participant 7, 22 and 34) were excluded because their performance during assessment was more than 3 standard deviations (4.3, 4.2 and 10.1, respectively) away from other participants' performance. Consequently, the analysis of preoperative simulation is based on twenty-nine participants.

A repeated-measures ANOVA, adjusted as Mauchly's test indicated the assumption of sphericity had been violated ($\chi^2 = 29.8$, $p < 0.001$) using Green-house-Geisser estimates of sphericity ($\epsilon = 0.638$), did not show a main effect of preoperative simulation on performance $F(2.55, 71.48) = 1.37$, $p = 0.261$, $\eta_p^2 = 0.047$. Although previously

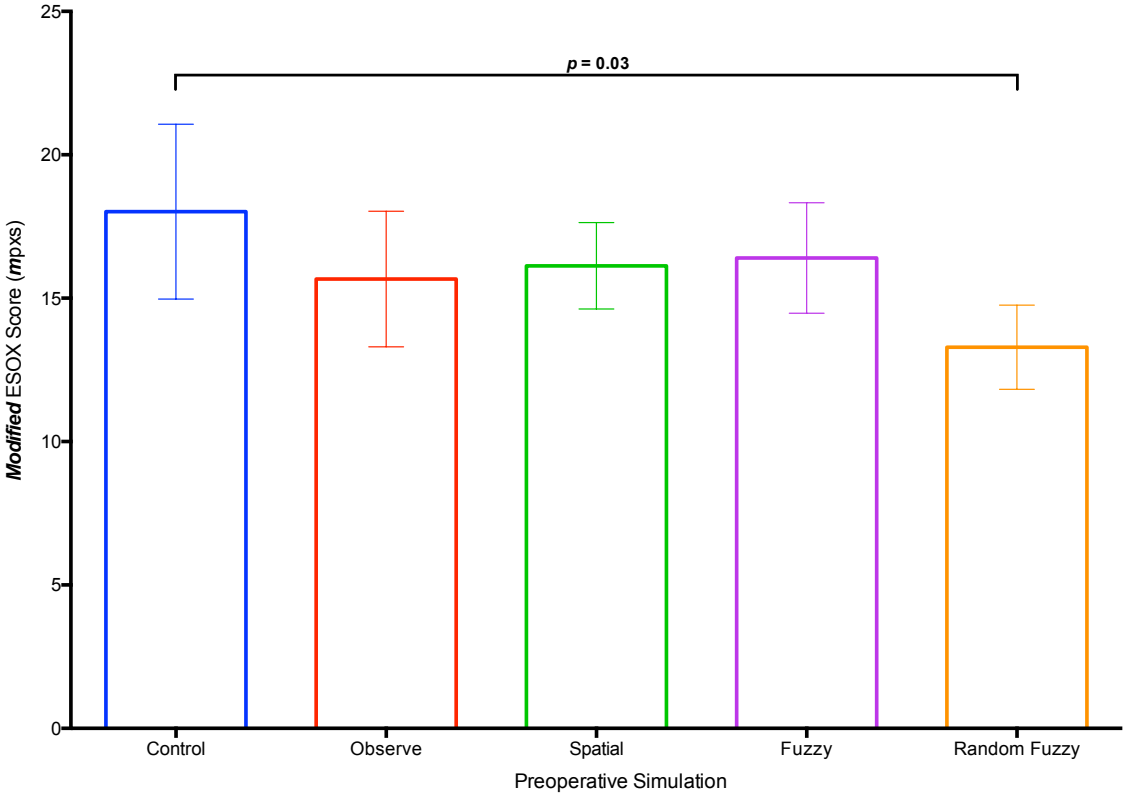


Fig. 9.2 Performance following Preoperative Simulation during Experimentγ. Significant effects are highlighted. The error bars represent the 95% CI of the mean.
Note: Smaller scores represent better performance.

described as post-hoc analysis, each experiment (α, β and γ) has planned comparisons - the interaction of interest is control *cf.* preoperative simulation. Consequently, while the overall ANOVA did not reach significance, it is not always necessary for the null hypothesis of homogeneity to be rejected before performing further comparisons [153]. Planned comparisons revealed significantly better performance following a colour, kinematic and *random fuzzy* spatial preoperative simulation, when compared to control ($p = 0.027$); Figure 9.2. Details of all comparisons are given in Table 9.1.

Table 9.1 All planned comparisons performed in Experimentγ.

Comparisons		Mean Difference	p
Factor 1	Factor 2		
Control	Observe	2.032	0.304
	Spatial	1.758	0.373
	Fuzzy	1.068	0.588
	Random Fuzzy	4.403	0.027

9.3 Discussion

The results of Experimentγ suggest preoperative simulation can have an effect on subsequent performance, but this improvement is seen in only one of the conditions - γPS:C,K&RFS. The nature of the effective preoperative simulation differed in Experimentγ from the previous two experiments.

Experimentα did not differentiate between colour and spatial information in terms of importance - the task could be completed using colour alone, spatial alone or a combination of both. However, participants showed a strong preference for utilising spatial information, both during the learning trials and at assessment of preoperative simulation, where spatial information alone was the most effective preoperative simulation method. Repetition of the same task was also a beneficial form of PS, and interestingly, a colour only preoperative simulation seemed (if anything) to hamper performance.

During Experimentβ colour information had primacy - given the random spatial arrangement each time the task was performed, spatial information was not helpful in completing the task. Providing a participant with the relevant information (colour sequence information) by way of watching a video was the most effective preoperative simulation. Requiring the participant to plan and execute movements that were not subsequently associated with the assessed task, in addition to this colour information (βPS:C), was not beneficial. Again, repetition of the task was also found to be an effective preoperative simulation.

Experimentγ was an attempt to balance the importance of spatial and colour information and create a hybrid of the two previous experiments. Consequently, it followed a similar design to Experimentα, but with a reduction in the reliability of spatial information - using fuzzy spatial location parameters results in a general spatial location that a square could occupy, but this location could change to a certain degree with each repetition. As well as balancing the informational importance between Experimentα and β, this reduction in spatial accuracy was designed to reflect clinical practice, as previously described. The altered distribution of importance between spatial and colour information in Experimentγ resulted in findings that are quire different to the preceding experiments.

Repeating a task prior to performing an assessed version of the same task was not beneficial during Experimentγ. Providing participants with colour and spatial information, by watching a video of performance (γPS:O), was not a beneficial form of preoperative simulation. The PS that *was* helpful during Experimentγ was arguably the most demanding of all the preoperative simulation tasks; γPS:C,K&RFS. One potential explanation for the observed results could be the ambiguity, introduced by

the inherent variability, of the Experimentγ task.

In all of the experimental tasks ($\alpha, \beta\&\gamma$) the stimuli of interest is considered bivalent; it has two dimensions (colour and spatial location), which is of relevance to the task being performed. Bivalent stimuli have dual affordances and have the potential to activate both the correct and incorrect response, in a bottom-up manner, which can lead to interference between the two competing responses [154]. Though first demonstrated ninety years ago [155], there has been increasing evidence that even single trials are enough to link a specific stimulus (e.g. picture-word combinations) to a specific task set (e.g. picture naming) and that later involvement of the same stimulus in a different task (e.g. word reading) is associated with significant slowing of the task, above and beyond that associated with switching between tasks [156–158]. The use of bivalent stimuli can create task ambiguity by the bottom-up cuing of the incorrect task set, and this ambiguity has a greater influence under conditions of task uncertainty [159]. It should also be noted that when utilising bivalent stimuli, participants also have the ability to filter out irrelevant variation in the stimulus, though this has been associated with a performance cost in some studies [160, 161], but not others [159].

During the performance of Experimentα, colour information was not utilised by participants in order to complete the task and during Experimentβ spatial information was unhelpful. Participants seem able to discount the irrelevant information in each trial, after they have been cued by an appropriate preoperative simulation, without a significant performance cost. However, Experimentγ was designed so that both colour and spatial information need to be used to complete the task optimally, meaning neither aspect of the bivalent stimuli can be simply discarded. Cueing participants to do this (as with γPS:O or γPS:C,K&S) could result in the stimulus being bound to the wrong response and consequently interfering with performance. The difference in the experimental tasks could be responsible for the difference in effective preoperative simulations between Experiments α, β and γ, as different information (colour, spatial or both) was relevant to performing the task. Alternatively, γPS:O may not be effective at engaging the colour set during Experimentγ, in contrast to Experimentβ, as discussed below.

The ineffectiveness of γPS:C,K&FS (the other contrast to Experiments α and β) may be explained by the lack of the reverse Stroop effect. A counterintuitive finding of recent experimentation in the cognitive psychology literature has been that it can be more difficult to switch to an easy task than a more challenging one. This was initially demonstrated by Allport, Styles and Hsich [157], and has been subsequently replicated by a number of studies [162–166].

Participants take longer to name a colour if displayed as an incongruent stimu-

lus (e.g. the word 'red' written in green ink) than for congruent ('green' written in green ink) or neutral ('xyz' displayed in green ink) stimuli - the Stroop effect [167]. However, word-naming trials are only minimally affected by displaying the word in a contradictory colour; there is only a marginal or no reverse Stroop impediment. This asymmetrical finding has been hypothesised to be due to the much greater practice a participant has at reading than colour-naming, resulting in a much greater tendency to name a word than the colour it is written in [168, 169].

Allport *et al.* [157] reasoned that when participants want to name a colour, a strong top-down control is required to overcome the formidable tendency to name the word and assist the weaker response to name the colour. This bias will necessitate greater activation of the colour-naming set and/or active suppression of word-naming set. It is thought these biases can persist over time, resulting in what Allport *et al.* described as task-set inertia (TSI). This results in persistent facilitation of colour-naming and/or suppression of word-naming even after a switch to a word-naming task. This causes a significant cost when participants switch from colour- to word-naming. Conversely, the more dominant word-naming task does not require a strongly imposed control as there is minimal competition to name the colour. This lack of strong control inertia makes is relatively easier to switch from word- to colour-naming.

Returning to Experimentγ, it is highly likely that participants will preferentially locate a square due to its spatial location, as humans are much more practiced at identifying objects by their location (*cf.* an arbitrary colour) [128–131], and as evidenced by the results of Experimentα. Consequently, utilising colour information is likely to require a strong top-down control to overcome the tendency to utilise spatial location. This can occur, as shown in Experimentβ. However, it is likely to result in task-set inertia. While Experimentγ was designed to create a balance between spatial location and colour information, even if this balance was perfectly 50:50, the innate preference for spatial information is likely to drive participants to try and utilise spatial information. Consequently, while γPS:O was initially envisioned as the effective colour-PS, as participants could also use it to glean spatial location information, they are likely to do just that. The same is true of γPS:C,K&S *and* γPS:C,K&FS - while spatial information is less reliable during γPS:C,K&FS when compared to γPS:C,K&S, it may still be reliable enough for participants to attempts to utilise the information, given their preference for spatial information. γPS:C,K&RFS presents the *least* reliable spatial information, as squares can move between ±1 and ±5 squares with each repetition. Under such conditions, the general spatial rules can be broken - using Figure 9.1a as an example, the orange square will remain in the bottom right corner for each repe-

tition of all PS conditions[1], with the exception of γPS:C,K&RFS, where, if able to move ± 5 squares, the orange square may end up in the bottom left or top right area of the array. This increased unreliability of spatial information may have been enough to encourage participants to utilise colour information during γPS:C,K&RFS. Such an approach would necessitate active facilitation of the colour-processing set and/or suppression of the spatial-location set, which would confer a TSI that could persist to the assessment trial. Use of colour information, by prior (and persistent) activation by the γPS:C,K&RFS preoperative simulation, during the assessment trial may explain why γPS:C,K&RFS improved participants' performance. Utilising colour information, being always consistent with each repetition of the task, may represent the most effective way of completing the Experimentγ assessment task, but is likely to be at odds with participants innate strategy preference. This suggests participants may not use the best approach to complete a task, unless primed by an appropriate preoperative simulation, despite previous exposure to the more effective method.

9.4 Implications for Preoperative Simulation in Clinical Practice

The results of the three experimental investigations of preoperative simulation present some fascinating insights into the potential use of PS in clinical practice.

Preoperative simulation has a significant effect on subsequent performance. All of the experimental conditions revealed some forms of PS that improved performance relative to control, and this tallies with the conclusions presented in Chapter 2.

A simplified, relevant preoperative simulation can prepare participants to complete a laparoscopic sequence learning task (Experiments α and β). The nature of this PS is dependent on the strategy utilised by the participant to complete the task, and may not require 'active' (motoric) performance of the participant.

Repetition of a task improves subsequent performance. The results of Experimentα and β demonstrate this phenomena, and this finding is supported by a wealth of literature from disparate fields. While the importance of this finding is arguably less relevant than the above statement, it does present a opportunity to examine the practice of PS in current clinical practice - see Chapter 10.

[1]During γPS:C,K&FS and γPS:O squares could move ± 3 squares, which resulted in squares remaining in the same general area, though their precise location changes with each repetition. During γPS:C,K&S the spatial location of squares was fixed.

Experimentγ illustrates what is potentially a very interesting finding; preoperative simulation has the ability to change a participants method of completing a task, resulting in improved performance. It should be noted that, during all of the experimental conditions, participants had to complete each repetition of the task before moving onto the next. Therefore, the assessment of preoperative simulation does not examine if participants could complete the task (*all* of the assessment tasks were completed), but if they could complete it optimally. Applying this to clinical practice, it may be that a surgeon could perform an operation, but in a way that is suboptimal. Preoperative simulation may be able to optimise a surgeons performance, even overcoming a preferred (but less optimal) strategic approach. This could be particularly influential for trainees, facilitating optimal performance in practice and the development of ideal operative strategies.

To try and extrapolate these experimental findings to current clinical practice, an investigation into the effect of operating list order, used as a real-world surrogate of preoperative simulation, on operative outcomes was conducted; Chapter 10.

Chapter 10

Evaluation of Preoperative Simulation in Clinical Practice

The experimental results of the previous chapters demonstrate that a simplified, relevant preoperative simulation can prepare a participant to complete a laparoscopic sequence learning task. These findings suggest that performing a relevant, simplified PS prior to performing laparoscopic surgery may improve a surgeon's performance and hopefully, by extension, patient outcomes. The findings of experiments α, β and γ are strengthened by the rigorously controlled nature of the investigations and are supported by the existing (less well controlled) literature and theoretical framework. However, in order to achieve adequate control of extraneous factors that may affect performance, the aforementioned experiments are significantly removed from the reality of performing minimally-invasive surgery. To counter this reduction in direct clinical relevance, a 'big data' investigation of the effect of preoperative simulation on actual clinical practice was utilised.

'Big data' is variously defined [170], but practically relates to large datasets that are routinely collected in healthcare and require advanced data processing and analytics. The utility of big data to surgical practice is being increasingly explored [171–174], and there is great hope for its application - it has been argued that it has the potential to allow personalised, precision healthcare [174, 175], and deliver significant cost savings [176]. However, the application of big data is in its infancy [177], and implementation issues remain [178]. Despite these concerns, if the problems surrounding implementation can be overcome, big data analysis remains a powerful method of examining current practice.

In order to investigate the effect of preoperative simulation in clinical practice, the natural experiment [179] of repeating a procedure during an operating list was

examined. It was reasoned that, if a surgeon performs more than one of the same type of procedure during an operating list (which frequently occurs in clinical practice), it may be that the first operation can act as a form of preoperative simulation and prepare the surgeon for the second case. This would mirror the findings of Experimentα and Experimentβ, both of which demonstrated performing a preoperative simulation which was the same as the assessed task significantly performed performance: §5.2, §7.2. Due to the inherent difference in people and their disease processes, repeating the same operation will not involve the surgeon performing *exactly* the same procedure, however, the same overarching sequence of steps that will need to be performed will be the same, and consequently repetition may aid performance. Alternatively, allowing a degree of exploration within a single procedure, by performing the same procedure on two different individuals sequentially, may improve performance during the second procedure (as per Experimentγ).

10.1 Preliminary Analysis

A preliminary, proof-of-concept investigation was conducted using locally available data. Utilising a previously published dataset of all laparoscopic cholecystectomies performed at Leeds Teaching Hospitals NHS Trust over a fifteen year period [178], we reanalysed the data, looking specifically at the effect of performing more than one laparoscopic cholecystectomy, by the same surgeon, on the same list.

Basic statistical analysis (Mann-Whitney U) was used to compare the operative time of the first laparoscopic cholecystectomy to that of subsequent laparoscopic cholecystectomies performed by the same surgeon on the same list. Operative time was shown to be significantly shorter for subsequent cases ($\tilde{x} = 62mins$) when compared to the first case ($\tilde{x} = 65mins$), $U = 676368$, $p = 0.0239$, $r = 13786.213$. Interestingly, despite shorter operating times, patients who were not operated on first had a significantly longer inpatient stay ($\tilde{x} = 1$) than the first case on the list ($\tilde{x} = 1$), $U = 783692$, $p = 0.0007$, $r = 15316.537$.

There are a number of potential explanations for the above findings, which cannot easily be differentiated between using the preliminary dataset. Subsequent operative times could be reduced because the surgeon was prepared to operate by the first case, while a longer inpatient stay was due to subsequent patients returning to the ward at a later time to those first on the operative list. Alternatively, surgeons may have operated faster, but not *better*, during subsequent cases and the longer inpatient stay was due to a higher complication rate experienced during subsequent cases. The longer

operative time found during the first case may be because some surgeons will prioritise what is thought to be the most difficult case to the beginning of an operative list. The dataset was also unable to quantify the effect of training on the reported outcome metrics - the reduction in operative times of subsequent cases may reflect a learning process in junior surgeons. The preliminary dataset is unable to explore any potential causal relationship between the data, but it does suggest that operative list order has an effect on operative performance and patient outcomes, and consequently is worthy of further investigation.

10.2 Analysis of Spire Healthcare Operative Data

To further investigate the effect of operative list order on surgical performance, operative data from all Spire Healthcare hospitals was obtained following ethical approval from the Spire Research Ethics Committee (SRECSE01, 13-Feb-2015); Appendix H.

Although the majority of healthcare delivery in the United Kingdom is provided by the National Health Service (NHS), the private sector is used by 10-22% of the population, dependent on region [180, 181], and this dataset was chosen to address the above research question for three reasons. First, these data could be pooled across multiple hospitals (a considerable logistical challenge in the NHS, due to the (quite public) failure to develop a single, integrated electronic patient record [182–184]). Second, all operations performed in private hospitals in the UK must be conducted by consultant (attending) surgeons - thus eliminating the effects of training and ensuring all procedures on a theatre list are performed by a single practitioner. Third, the use of private healthcare data from the UK more closely mirrors that of other Western countries [185].

Data from all operations performed across Spire Healthcare's thirty-eight UK hospitals between 1st April 2013 and 31st May 2015 was extracted from Spire Healthcare's electronic theatre management system (SAP SE, Walldorf, Germany). Patient demographics, procedural/operative information, prognosticators of operative outcome (age and American Society of Anaesthesiologists' classification of physical status; ASA-PS [186]) and hospital length of stay (LOS) were included in the dataset. Age was divided into blocks, to allow adequate anonymisation of data (<18 years; 19-24; 25-34; 35-44; 45-54; 55-64; 65-75 and >75 years). Only cases with complete information were included. This dataset contained 478,713 operations.

The 35 most frequently observed operations in the dataset were the primary focus of investigation. No restriction was placed on the surgical sub-speciality, the type of

Table 10.1 Illustration of absolute and procedure specific list number and 'switch' classification

Procedure	Absolute List No.	Procedure Specific List No.	Switch?
Laparoscopic Cholecystectomy	1	1	
Open Inguinal Hernia Repair	2	1	Yes
Laparoscopic Cholecystectomy	3	2	Yes
Laparoscopic Cholecystectomy	4	3	
Open Inguinal Hernia Repair	5	2	Yes

procedure performed, or the techniques used by the operating surgeon to perform the procedure. The collated data were parsed to allow further analysis; individual surgeon's operating lists were identified and any instance when one of the most frequent 35 operations was performed during an operating list was included in the dataset (98,291 lists). Any other procedure performed during said list was also included in the dataset (total procedures = 255,757 cases).

Component operations were allocated 'absolute' and 'procedure specific' order numbers. The absolute list number refers to the number of procedures performed by the operating surgeon on the list, whereas the procedure specific list order is the number of times a certain procedure has been performed by the surgeon on a list. All cases that did not involve repeating the same procedure were coded as a 'switch' – because they involved some form of task switching [187]: Table 10.1. Procedures were also classified by modality (open or minimally invasive surgery [MIS]) and complexity, as per the AXA Specialist Procedure Codes, which is used to grade the magnitude of surgical procedures in UK independent hospitals [188]: Table 10.2.

From the original dataset comprising 478,519 individual procedures, we excluded 8,807 cases because they had no surgeon ID associated with them and 1,422 cases where no start time was recorded. Thirty-two duplicate records were also removed (Figure 10.1). Whilst such instances were relatively trivial to identify, a more difficult challenge in analysing routinely collected data lies in identifying cases where erroneous data might have been entered e.g. the wrong start type or procedure type or instances where missing data might have been due to procedures that were ultimately cancelled. All of these factors are likely to influence procedure order classification. This introduced noise but we reasoned that the noise would work against the hypothesis being tested (because our hypothesis suggests that the preceding operation n–1 affects the subsequent one, so in instances where data are missing, using n-2 would make it more likely that we would reject the hypothesis). Importantly, because of the statistical power afforded to us by a dataset of this size, we were willing to tolerate this noise in the data and adopt a conservative approach to the hypothesis testing and therefore did not adjust

Table 10.2 Procedure modality and complexity classification of the included 35 operations

Procedure Code	Procedure Description	Modality	Complexity
G451	Oesophago-gastro-duodenoscopy (OGD, with biopsy of lesion)	MIS	Intermediate
H229	Endoscopic Examination of Colon (Unspecified, Diagnostic)	MIS	Intermediate
F0910	Extraction of Impacted/Buried Tooth/Teeth	Open	Intermediate
O291	Subacromial Decompression	Open	Intermediate
T202	Primary Inguinal Hernia Repair (with mesh)	Open	Intermediate
J183	Cholecystectomy	Open	Complex
V544	Spinal Injection	NA	Intermediate
B3121	Bilateral Augmentation Mammoplasty	Open	Intermediate
Q1800	Hysteroscopy	MIS	Intermediate
Q3800	Laparoscopy and Therapeutic Procedure (Gynecological)	MIS	Major
A5770	Facet Joint Injection	NA	Intermediate
C751	Lens Implant	Open	Major
A577	Injection Around Spinal Nerve Root (Therapeutic)	NA	Intermediate
J1830	Laparoscopic Cholecystectomy	MIS	Complex
W371	Primary Total Hip Replacement with Cement	Open	Complex
W401	Primary Total Knee Replacement (with cement)	Open	Complex
T2000	Primary Inguinal Hernia Repair	Open	Intermediate
W903	Joint injection (Therapeutic)	NA	Minor
W381	Primary Total Hip Replacement (without cement)	Open	Complex
A6510	Endoscopic Carpel Tunnel Release	MIS	Intermediate
W822	Endoscopic Resection of Semilunar Cartilage	MIS	Intermediate
W4210	Total Knee Replacement +/- Cement	Open	Complex
W3712	Primary Total Hip Replacement +/- Cement	Open	Complex
W8500	Knee Arthroscopy (multiple)	MIS	Major
A5210	Epidural Injection	NA	Minor
C7122	Phakoemulsification of Lens with Implant	Open	Intermediate
A651	Carpal Tunnel Release	Open	Intermediate
W8200	Arthroscopic Meniscectomy	MIS	Major
M4510	Cystoscopy (Diagnostic)	MIS	Minor
H2002	Colonoscopy (Diagnostic)	MIS	Intermediate
G6500	OGD (Diagnostic)	MIS	Minor
G8082	OGD + Colonoscopy	MIS	Intermediate
H2502	Flexible Sigmoidoscopy (Diagnostic)	MIS	Minor
W9030	Joint Injection (with image guidance)	NA	Minor
25120	Dorsal Root Ganglion Block	NA	Intermediate

list order numbering (which would have required us to make subjective inferences).

Operating time was employed as the primary outcome measure. Generally, this measure was defined as the time from skin incision to skin closure and in procedures where a skin incision is not made (such as endoscopic examinations), we analysed the time taken for the procedure to be performed (defined as the time from insertion to withdrawal of the endoscope in the case of endoscopic examinations). This was chosen because: (i) it is strongly correlated with surgeon performance (the focus of this study) [189] and (ii) others have shown a relationship between this metric and clinical outcomes across a range of operations [189–192]. In addition, operative time is routinely recorded in Spire Healthcare hospitals, and is not affected by a loss of patients to follow up, unlike other measures of clinical outcome e.g. hospital mortality. Length of hospital stay (in minutes) was investigated as a secondary outcome measure.

10.2.1 Statistical Analysis

Operation times are zero-bound and present a skewed distribution [193]. Therefore, for all analyses, a change in natural logarithmic operation time (which can be seen as equivalent to measuring proportional time changes for relatively small magnitudes) was focussed on. A model was created to capture the effects of absolute list order, procedure specific list order and procedure switching across the full range of list positions to understand the relationship between list composition and operating time. As different operations might yield distinctly different patterns of results, analysis was conducted at an operation level to allow individual operations to be compared against the same types of procedures (i.e. total knee replacements were only compared to total knee replacements and not to lens phakoemulsification).

The dataset included information on factors known to correlate with postoperative outcomes (ASA-PS [194–197] and age [198–200]) and consequently these potential confounds were controlled for. Different patient ages and ASA-PS scores along with different surgical procedures would imply different 'normal' operation lengths and therefore, these baseline operation lengths were treated as random effects, shared by all operations of the same type, on the same age block and with the same ASA-PS. In reality, operations on similar age blocks or similar ASA-PS will have similar baselines: For example, a 34 year old with a ASA-PS classification score of II is more similar to a 40 year old with a score of II than an 80 year old with a score of IV. However, we did not wish to make assumptions about the relationship between operating time and these factors. Instead, we adopted a statistically more conservative approach by assuming that these random effects are independent between pairs of operations (unless all three

of these variables were identical). The restricted-likelihood maximisation via the Lme4 package [201] was used to fit the linear mixed effects model in R (R Development Core Team, 2016) and estimate the effect-size and probability values (alpha threshold of $p < 0.05$) for the fixed effects of list order (absolute and procedure specific) and switching.

For a closer examination of the primary effects observed in the data, a form of matched analysis was performed on a subset of the data. This analysis was inspired by (but not identical to) a novel method for identifying causal relationships in natural experiments [202]. Here, the data were stratified into multiple sets of pairs by explicitly matching individuals who had the same age, ASA-PS and operation type, but differed in list order by one position. Specifically, the data was initially filtered by procedure type, then all cases that were ordered as procedures 1 and 2 were identified and separated into different data frames (List Order 1 and List Order 2). Then all cases in List Order 1 (presented in a randomly determined order) were sequentially cycled through to compare with elements of List Order 2 which had the same Age block and ASA-PS score. If a case could be matched, then this pair was included in the subsequent analysis and removed from the pool. In the case of multiple matches from List Order 2 with List Order 1, the computer programme randomly selected one case for the pair and the non-selected case(s) were returned to the pool for a future possible matches. Each patient was paired to only one other individual and only patients for whom a pair could be found were included. The matching process terminated when no more unique pairs could be found. This approach presents a method for statistically controlling for all the potential confounding variables available in our dataset.

In addition to these primary analyses, an investigation was undertaken to identify: (i) whether these effects translated across modality (i.e. open and MIS); and (ii) if the impact of list order was modulated by procedure difficulty. For this first question, procedures were separated by classifying them as those performed using open techniques and MIS and for the second question, procedures were classified by difficulty and subsequently the matched analysis was repeated.

To provide a measure of the magnitude (or effect size) of the analysed variables on list order, the change in the log scale for the linear effects model and mean difference in the log duration of the procedures in the matched analysis is reported. Change in log duration is, to a high degree of approximation, the geometric average of the proportional percentage (%) change in operation duration, thus change % for all outcomes is referred to, to provide an intuitive means of understanding these data. The Flow Chart in Figure 10.1 provides an illustration of how sample sizes for the linear effects model and matched analyses were determined.

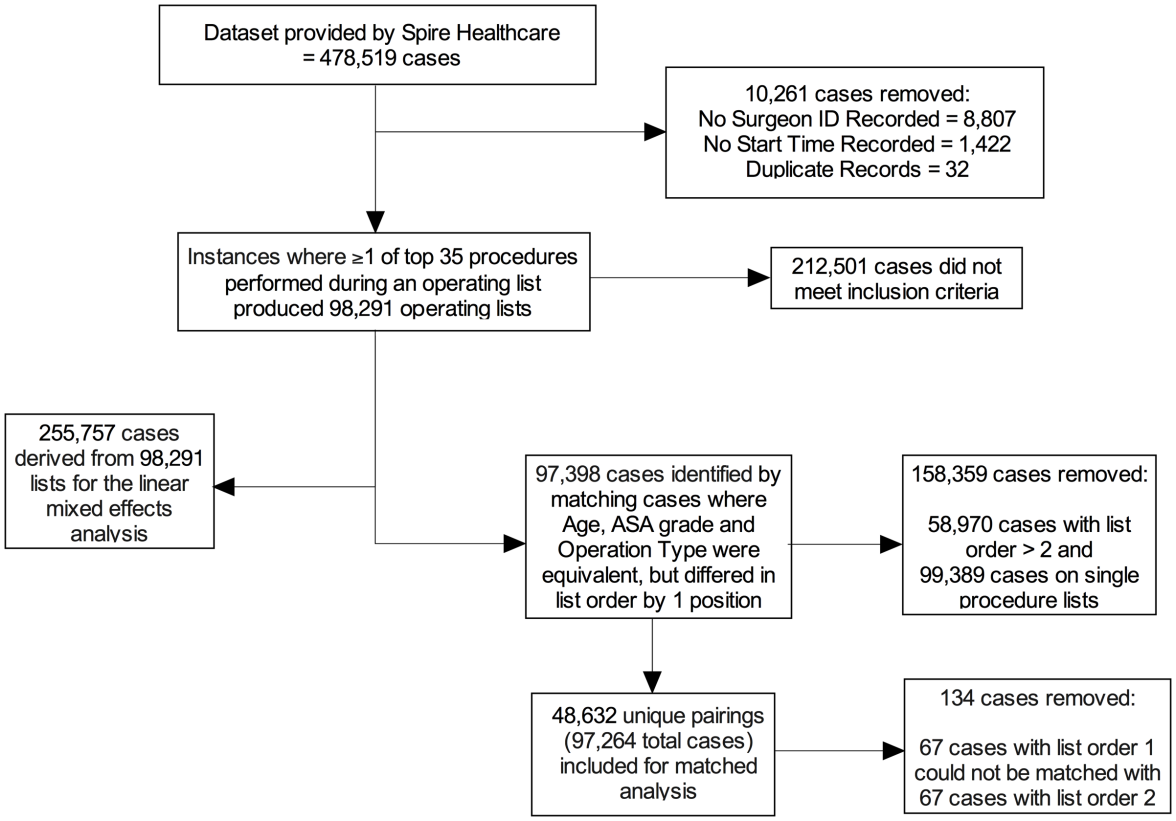


Fig. 10.1 Flow chart illustrating how sample sizes were determined for the linear mixed effects and matched analyses from the original dataset.

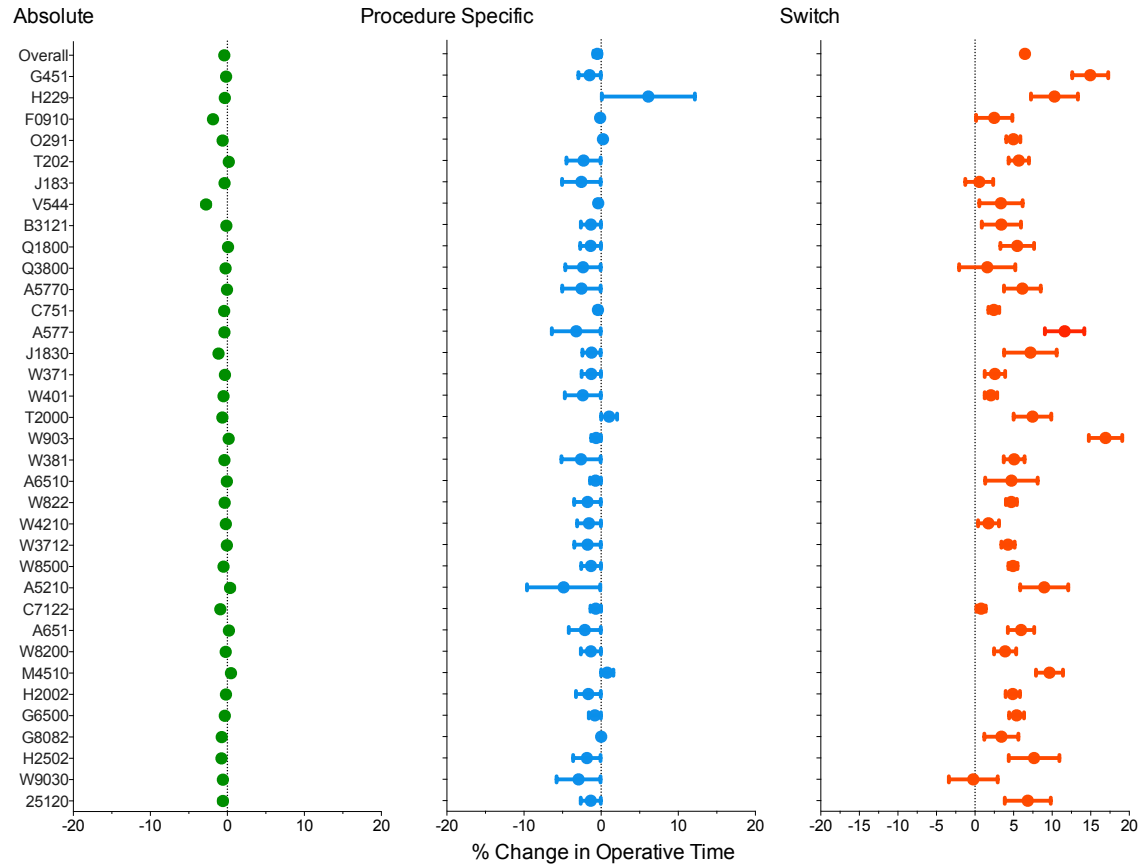


Fig. 10.2 Forest plot showing the percentage change in operative time for the influence of fixed model parameters absolute list order, procedure specific list order, and procedure switch. Negative values indicate the percentage reduction in operative time given an increase in each parameter and positive values indicate the percentage increase in operating duration. The top row in each panel shows the overall effect of each fixed parameter. Error bars denote standard error.
Note: For details of procedure codes see Table 10.2.

10.3 Results

The linear mixed effects model revealed statistically reliable differences in operating time for the fixed effects of absolute list order, procedure specific list order and switching ($p's < 1 \times 10^{-16}$) when pooling across operations. The effect sizes – which can be treated as percentage change in operative time as a function of list position change – were largest for procedure specific list order and switching. The percentage change in operative time for each procedure can be seen in Figure 10.2.

For absolute list order, statistically significant effects suggesting each position in the list decreases operation time by 0.39% (SE = ±0.02%) were found across all operations. These effects are substantially greater when the same procedure is repeated in a list, with the effect of procedure specific list order leading to a 0.98% (SE = ±0.05%)

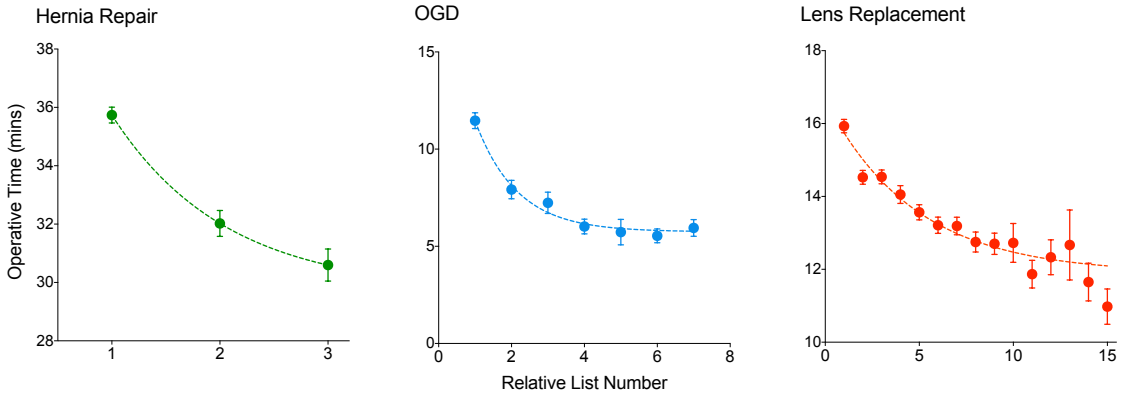


Fig. 10.3 The effect of procedure specific list order on operative time (minutes) for open hernia repair, oesophageal-gastro-duodenoscopy (OGD) and lens replacement. Error bars denote standard error.

reduction. A cost associated with switching between different operations in a list – leading to an increase in operation time by 6.48% (SE = ±0.22%) was also found for each increase in position on list order.

To illustrate this effect on individual procedures, Figure 10.3 demonstrates the effect of procedure specific list order on operative time for three exemplar procedures, routinely performed a differing numbers of times on an operative list. There was a marked similarity in the pattern of results across these distinct procedures, indicating that fatigue, inattention and monotony-related performance impairments following multiple repetitions of a procedure are not present in these data.

Using the same linear mixed model to analyse LOS, statistically reliable effects of absolute and procedure specific list order ($p's < 1 \times 10^{-16}$), but not switching procedure ($p = 0.0136$) were found. Specifically, the data indicates that for every increase in absolute position, LOS increases by 0.53% (SE = ±0.03%). However, procedure specific list order results in a decrease in LOS by 0.72% (SE = ±0.07%).

The matched analysis allows a focus on the effects of repeating a procedure in more detail on the primary outcome measure of operating duration. A total of 48,632 pairs were matched out of a maximum 48,699 cases (99.86% of total cases; with the sample size being constrained by the number of cases with procedure specific list order of 2 in the dataset). Here, we found statistically reliable improvements ($p < 0.05$) in 29 out of the 35 procedures, with changes in operative time ranging from a reduction of 3.84% (SE = ±1.21%) to 17.25% (SE = ±3.34%): Figure 10.4. Pooling across all 35 procedures showed a 6.18% (SE = ±0.27%) reduction in operating time ($p < 1 \times 10^{-16}$) on average when performing the second procedure relative to the first.

Supplementary analyses allowed these results to be teased apart in more detail. The

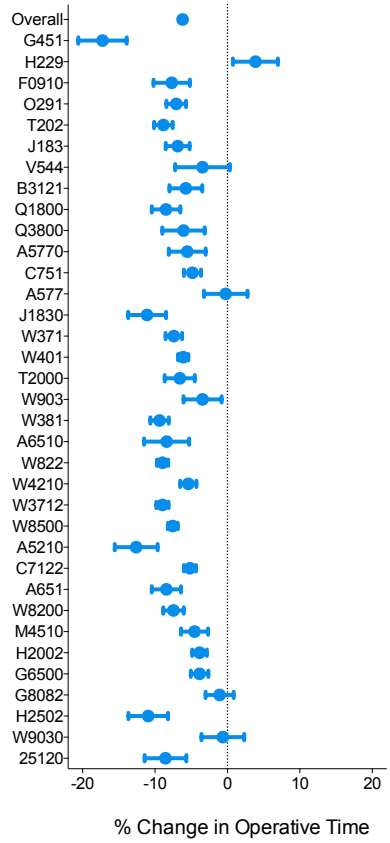


Fig. 10.4 Forest plot of matched analysis illustrates the percentage change in operative time for procedure specific list order 2 procedures compared to list order 1. Twenty-nine procedures showed statistically reliable reductions in operating time, with a pooled effect of 6.48% reduction. Error bars denote standard error.
Note: For details of procedure codes see Table 10.2.

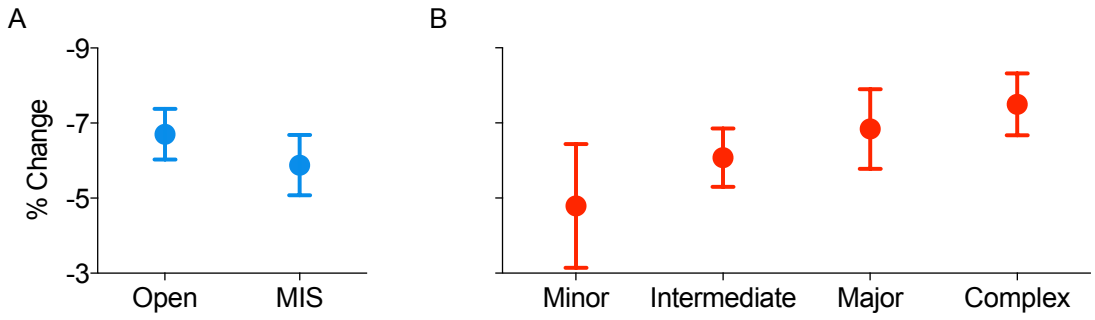


Fig. 10.5 Percentage change in operation time for (A) modality and (B) procedure complexity in matched analysis. There were no statistically reliable differences in each panel, illustrating the effects are consistent across modality and procedural complexity. Error bars denote standard error.

matched analysis was conducted separately for open and minimally invasive procedures, and comparable effects of list order on operating duration were found – indicating that this phenomenon transcends operating modality: Figure 10.5A. Finally, by separating procedures based on their complexity, a weak positive trend was found but, overall, there were no differences in effect size as a function of difficulty, with the reduction in operating time ranging between 4.80%- 7.50%: Figure 10.5B.

10.4 Discussion

The above data demonstrates, for the first time, that the order in which procedures are performed has a predictive relationship with operation duration. Most notably, this result was relatively consistent across thirty-five different types of operations, suggesting the effect is independent of procedure type. The effects were similar for open and MIS procedures, and operations of differing complexity yielded comparable results. In contrast, switching between different procedures resulted in increased operating duration. These changes in operating time are particularly significant given that they were observed in highly trained individuals with several years of practice (consultant surgeons). The results are all the more remarkable when one considers the wide range of factors that can potentially influence procedure duration. The consistency of this pattern of results across procedure type, modality and complexity provides compelling evidence that operating list order plays an important role in surgical performance.

The reduction in operative time may be a consequence of the attending surgeons being *prepared* to operate by the preceding case. This finding is supported by the more significant reduction in operative time demonstrated by increasing procedure specific list number (*cf.* absolute list number) and that switching tasks produced the oppo-

site effect. This suggests that the preceding case may act as a form of preoperative simulation. It would also suggest that there are generic aspects operating that can be facilitated by *any* preceding case (as demonstrated by reduced operative time with increased absolute list number), but the effect is greater with a *specific* form of preoperative simulation, as evidenced by the greater effect on operative time with increased procedure specific list number.

This hypothesis is further supported by the findings of list order on LOS. These initially seem contradictory; increasing absolute list number increases LOS, whereas increasing procedure specific list order reduces LOS. However, these results may also be explained if a *specific* preoperative simulation is more effective than a generic preoperative simulation. That absolute list number increased LOS seems intuitive, as patients who are operated on later in a list are less likely to be discharged on the same day as surgery (although Spire Healthcare hospitals will discharge patients up until 22:00 for day-case procedures, if a procedure hasn't finished until 21:30, it is less likely that said patient will be discharged the same day *cf.* a patient whose operation finished at 11:00). This effect is likely to act to increase the LOS for increasing list numbers. However, a specific preoperative simulation may have a profound effect on performance, resulting in better clinical outcomes and a reduced LOS. For procedure specific list order, this improved performance may exert a greater effect than the inherent increase in LOS associated with later cases, indicating a specific preoperative simulation is particularly impactful (and reducing LOS for increasing procedure specific list order). A generic preoperative simulation (as provided by increasing absolute list order) does not produce a large enough effect to overcome the inherent effect of increasing LOS on later procedures.

Switching procedures was associated with a cost. Increased task duration as a consequence of switching has long been established in experimental psychology [187], but this is the first demonstration of its influence on surgical performance. This finding is significant as the vast majority of the experimental psychology literature demonstrates this effect on dramatically simplified tasks (*cf.* surgery). It is also important to note that the cost of switching described in the experimental psychology literature is usually a time cost; increased errors are associated with switching, but the finding is less reproducible experimentally [187], and usually of a lesser magnitude, than an increase in time: see Figure 10.6 as an example. It is possible that participants take longer in order to ensure fewer errors occur, under certain conditions.

A switching cost may also explain some of the experimental findings of the ESOX experiments, for example, the effect of a colour preoperative simulation on performance during EXOSα (see Chapter 5). During ESOXα, spatial information was strongly

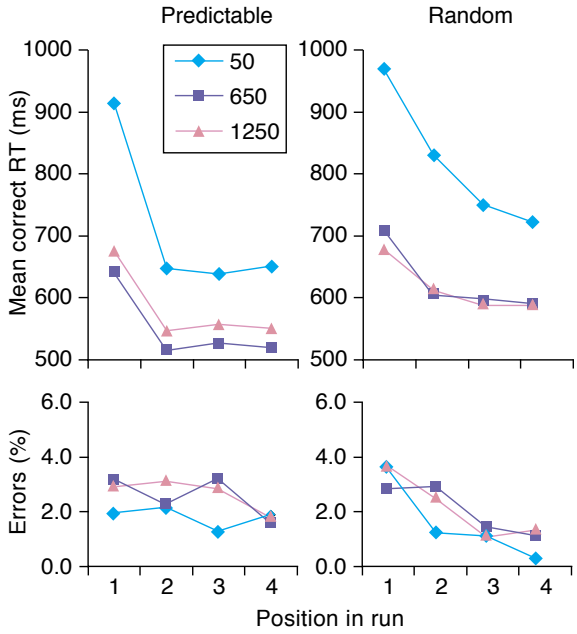


Fig. 10.6 The effect of predictable and random switching, reproduced from Monsell *et al.* [203] (Experiment 2). In this experiment, designed to look at the effect of predictable and unpredictable switching, the tasks were to classify the digit as either odd/even or high/low, with a left or right key-press. The effect of switching was significant in terms of time (RT) for both conditions, but only made a significant difference to errors when switching was random, not predictable.

preferred by participants, both during the learning and assessment of preoperative simulation phases. A colour preoperative simulation seemed to diminish performance (though this did not reach significance *cf.* control). Having undertaken the colour preoperative simulation (which was performed on a tablet computer), participants then had to *switch* to perform the experimental task in a laparoscopic box trainer, possibly reverting to try and utilise spatial information. This switch in approaches and/or methods of performing the task may be responsible for the lack of efficacy of a colour preoperative simulation during ESOXα.

Switching was associated with a significant increase in operative time, which tallies with the experimental psychology literature [187, 204, 205] and the previous experimental findings. Switching had no effect on LOS which may be because surgeons took additional time in order to ensure errors did not occur. Errors are inconsistently reported in clinical practice [206], but the majority of errors in surgery have been shown to be as result of technical errors during routine procedures performed by experienced surgeons [207], which can lead to significant disability and death [208]. Any significant technical error is consequently likely to increase LOS.

As well as demonstrating a cost associated with switching, the experimental psychology literature also shows this cost can be reduced if advanced knowledge of an upcoming task is provided [187], although preparation does not usually eliminate the cost entirely [209, 210]. The operating surgeon will know, in the elective setting, what procedure is due to be performed next and this knowledge, in conjunction with a prioritisation of error-free surgery at a additional time cost, may be responsible for the switching results demonstrated. The causative mechanism for the cost associated with switching is still in dispute; although initial theories attempted to explain the phenomena using a single description [157, 211], there now seems to be consensus that there are a multitude of causes, although the precise mix remains in contention [187].

One limitation of the current analyses is that it cannot speak to the exact mechanism of performance facilitation. Such a method of study as utilised here results in a loss of experimental control over the factors that affect surgical practice. However, there are significant advantages with a big data approach, which stem from the very large sample sizes involved. Some of the importance placed upon experimental control during other forms of investigation is due to the loss of statistical power that can result from uncontrolled measurement [212]. A loss of statistical power is not an issue with large sample sizes. Consideration does need to be given to the ways in which a loss of experimental control could introduce systematic confounds into the data. 'Slicing' can attempt to rectify this issue. During conventional experimentation, attempts are made to control potential confounds in advance of conducting the study. The slicing of big data allows the *post hoc* control of potential confounds, by selecting multiple different sub-datasets, without a loss of statistical power [212] (for example, the matched analysis performed on pairs of patients explicitly matched for the potential confounds of age and ASA-PS). Additionally, the results presented in this chapter triangulate well with existing empirical work showing preoperative simulation reduces operative times, as described in Chapter 2, and demonstrated experimentally in Chapters 5, 7 and 9.

The data presented here also have practical implications. An overall 6.2% saving in operation time (which could be as large as 18% in some procedures) was found for repeating the same procedure on the list – even after controlling for age and ASA-PS. This control is particularly important as anecdotal evidence indicates that surgeons typically take these factors into account when compiling their lists, but the data indicate that the process of list ordering itself impacts on operative time above and beyond the variance captured by age and ASA-PS.

It is possible to argue that these effects are modest on average (for example, performing two successive laparoscopic cholecystectomies results in the second procedure being 6 minutes shorter, on average). On the other hand, the cumulative accrual of

even modest gains can produce substantial benefits when scaled across health services - particularly important in the context of surgical service delivery given the growing economic pressures to optimise elective surgeries [213]. This idea is central within competitive sport where many disciplines have been transformed by identifying small but readily implementable changes. These 'marginal gains' have produced a substantial aggregate effect of performance improvements, as evidenced by ever-improving world records. While the full extent of the impact of these effects remains to be seen, to put the results in context, it is worth noting the potential benefits. Just over half of the cases (52%) analysed are a switch from the preceding case, and the results demonstrate switching is responsible for a 6.48% increase in operative time. Consequently, as a crude approximation, in Spire Leeds Hospital over the course of a year, switching between different operations accounts for approximately 24 days of operating time.

In summary, the results demonstrate, for the first time, that the duration of an operation is modulated by list order and this effect is robust to operation complexity and modality.

Chapter 11

Conclusions

In this thesis, three approaches to investigate the effect of preparing to operate before performing minimally-invasive surgery have been undertaken; systematic review, experimental psychology, and big data analytics.

Having outlined the theoretical background for this work (Chapter 1) and performed a systematic review of current understanding (Chapter 2), a bespoke simulator was developed to study the effect of preoperative simulation under controlled laboratory conditions (Chapter 3). The following six chapters (4-9) present the results of three controlled, randomised crossover trials examining the effects of preoperative simulation on a laparoscopic sequence learning task. Finally, a big data investigation of the effect of a proxy of preoperative simulation in current clinical practice was presented (Chapter 10).

The main results of this thesis are summarised in greater detail in §11.1. Following this, in §11.2, is a discussion of some of the main areas that have emerged for further work.

11.1 Summary of Results

Chapter 2 presents a review of the current literature evidence for preoperative simulation, conducted in accordance with PRISMA guidelines [69]. From four hundred and eighty-three articles identified using the search strategy, thirteen were included for review. A variety of experimental designs were included, from RCTs to case studies that did not contain a control group. All of the included studies suffered from methodological shortcomings, but twelves of the thirteen studies concluded that preoperative simulation improves surgical performance, in both real-patient and simulated environments. The work in this chapter is published in Ref. [46].

In stark contrast to the development of novel pharmaceutical compounds, the majority of surgical techniques and procedures have not been developed in a structured manner, utilising animal or artificial models, prior to their practice on patients [214]; they have evolved in current clinical practice, through a process of 'trial and error' [215]. Though it has been argued that such a process drives innovation [216], it also presents an ethical conundrum by experimenting on patients. Chapter 3 attempts to address such ethical concerns by detailing the development of a bespoke laparoscopic sequence learning task that can be utilised in the assessment of preoperative simulation under controlled, laboratory conditions, building on some of the theoretical frameworks discussed in Chapter 1. Though somewhat necessarily removed from the practice of minimally-invasive surgery, the task (ESOX) allows for detailed metrics of performance to be generated without (significant) risk.

Utilising the ESOX program, three randomised, controlled crossover experiments were conducted, Experiments α, β and γ. All shared similar experimental designs of a learning phase (Chapters 4, 6 and 8), designed to teach participants the specific laparoscopic sequence learning task and a subsequent assessment of the effect of preoperative simulation on the previously learnt task (Chapters 5, 7 and 9).

The experimental chapters demonstrated that repetition of a procedure improves subsequent performance; a theorem that was subsequently demonstrated in clinical practice in Chapter 10. Perhaps more relevant to the concept of preoperative simulation, a simplified, relevant procedure was also shown to be able to improve subsequent performance. Exactly what form the preoperative simulation should take is determined by the nature of the task/operation being performed and is likely to reflect how that procedure was learnt. Interestingly, preoperative simulation was also shown to be able to modify participants' approach to completing a task, overriding their initial preferred method, to condition participants to use a *better* method of completing the task. These findings could have the potential to improve trainee performance, allowing a shortening of the learning curve and a hastening of independent practice (of particular import due to the reduction in both training time and opportunities afforded to current trainees [217]). Additionally, it may have a significant impact on expert performance by mitigating the effect of switching operations, demonstrated in Chapter 10, which cannot always be avoided in practice. The majority of work in this chapter is published in Ref. [218].

Finally, in Chapter 10, the results of a big data investigation into the effects of a proxy of preoperative simulation in current practice was presented. As there are currently no uniformly accepted methods of preoperative simulation and methods of preparing to operate vary significantly between surgeons, the natural experiment of

repeating a procedure during a operating list was used to investigate PS, with the first case acting as the preoperative simulation for the second. For the first time, the order in which procedures are performed was shown to have a predictive relationship with operative duration (a surrogate for operative quality). This finding was relatively consistent across thirty-five different procedures (open and MIS, and from minor to complex), suggesting the effect is independent of procedure type. This finding was reinforced by the demonstration of the effect of switching procedures, which lead to significantly increased operative times. The work in this chapter has been published in Ref. [219].

11.2 Further Work

There are two main further areas of work emerging from this thesis.

The potential of preoperative simulation has been demonstrated in the literature, through empirical laboratory-based experimentation and by big data investigation of current practice. Expanding on this evidence base by using preoperative simulation in simulated clinical practice is the next step in developing this nascent field. The most effective way of proceeding might be to incorporate preoperative simulation into current training; this could both accelerate the learning curve of trainees and demonstrate the relevant information that may form the basis of an effective preoperative simulation.

The utility of big data investigation has been effectively demonstrated in this thesis. Further exploration of large-scale datasets could provide further insights, particularly by utilising NHS data, which may allow parsing of the effect of preoperative simulation on training, among other avenues of investigation.

Appendix A

Risk of Bias Tables

Risk of bias tables for included RCTs, as per Cochrane risk of bias tool [70].
Risk of bias table for included cross-over trials, as per Mills *et al.* [71].

Table A.1 Summary of risk of the included RCTs

	Random Sequence Generation	Allocation Concealment	Bliniding of Outcome Assessment	Incomplete Outcome Data Addressed	Selective Reporting
Plerhoples *et al.*	?	?	+	-	+
Weston *et al.*	+	+	+	+	+
Chen *et al.*	+	+	-	+	+
Lendvay *et al.*	+	+	+	+	+
Arora *et al.*	?	+	+	-	+

Table A.2 Risk of bias table for Plerhoples *et al.* [74]

	Judgement	Quote	Comment
Random Sequence Generation	Unclear Risk	"Laparoscopic novice were stratified by training level and then randomized"	Participants stratified, but no description of randomisation method
Allocation Concealment	Unclear Risk	"Laparoscopic novice were stratified by training level and then randomized"	Participants stratified, but no description of randomization method
Blinding of outcome assessment	Low Risk	"Assessment performed using... ProMIS"	Performed by simulator
Incomplete outcome data addressed	High risk		Only percentages reported
Selective Reporting	Low Risk	"Trend towards improvement on all measured metrics"	Results given for all described metrics

Table A.3 Risk of bias table for Weston *et al.* [72]

	Judgement	Quote	Comment
Random Sequence Generation	Low Risk	"Randomization... based on random number sequence"	Probably done
Allocation Concealment	Low Risk	"Randomization... coordinated by clinical trial officer"	Probably done
Blinding of outcome assessment	Low Risk	"Blinding... coordinated by clinical trial officer"	Probably done
Incomplete outcome data addressed	Low Risk	"113 patients consented... which provided 75 complete digital recordings"	Missing outcome data balanced in numbers across intervention groups
Selective Reporting	Low Risk		All listed outcome measures reported

Table A.4 Risk of bias table for Chen *et al.* [73]

	Judgement	Quote	Comment
Random Sequence Generation	Low Risk	"SPSS... used to randomize warm-up assignment"	Probably done
Allocation Concealment	Low Risk	"Residents were given sealed envelopes of the warm-up assignments"	Probably done
Blinding of outcome assessment	High Risk		Participants assessed by attending surgeon after case
Incomplete outcome data addressed	Low Risk	"184...cases were eligible for randomization... Complete data was available for 91"	Missing outcome data balanced in numbers across intervention groups
Selective Reporting	Low Risk		All listed outcome measures reported

Table A.5 Risk of bias table for Lendavay *et al.* [75]

	Judgement	Quote	Comment
Random Sequence Generation	Low Risk	"Permuted blocks randomization was used."	Probably done
Allocation Concealment	Low Risk	"Randomization assignments were provided to each site in sealed, fully opaque envelopes"	Probably done
Blinding of outcome assessment	Low Risk	"Errors... were documented in real-time by study personnel"	No description of blinding
Incomplete outcome data addressed	Low Risk		No missing outcome data
Selective Reporting	Low Risk		All listed outcome measures reported

Table A.6 Risk of bias table for Arora *et al.* [84]

	Judgement	Quote	Comment
Random Sequence Generation	Unclear Risk	"A randomized controlled study design was utilized"	No description of randomization method
Allocation Concealment	Low Risk	"... participants were then randomly assigned using the closed envelope technique..."	Probably done
Blinding of outcome assessment	Low Risk	"[Performances]... were video recorded and blindly assessed by 2-experienced laparoscopic surgeons and trained raters "	Probably done
Incomplete outcome data addressed	High Risk		Only summaries of outcome data reported
Selective Reporting	Low Risk		All listed outcome measures reported

Table A.7 Summary of risk of the included cross-over trials [76–80]

	Lee *et al.*	Calatayud *et al.*	Moldovanu *et al.*	Kroft *et al.*	Willaert *et al.*
AB / BA Design	Yes	Yes	Not clear	Yes	Other
Carryover concept recognised in the methods	Yes	Yes	No	Yes	Yes
Washout	Used	Used	Not mentioned	Used	Used
Randomisation	Not explicit	Explicit	Not clear	Explicit	not clear
Allocation concealment	No	Yes	No	Yes	No
Sample size calculation	No	No	No	No	No
All participants accounted for	Yes	Yes	Yes	Yes	Yes
Conclusions based on	Between groups	Between groups	Between groups & within-participant	Between groups	Between groups

Appendix B

Outcome Measures of Included Studies

Table B.1 Outcome measures reported by the included studies [72–84].

Authors	Outcome Measure	Details	Results	Significance
Kroft *et al.*	Timing Score	600 - Time - 10(penalties)	NS †	
Willaert *et al.*	Global Rating Scale	OSATS-derived Global Rating Scale	Patient-specific PS > Generic PS > control	$p = 0.038$ & $p = 0.050$
		Procedure-specific rating scales	NS	
	Dexterity Metrics'	Total Procedure Time	Patient-specific PS > Generic PS > control	$p = 0.001$
		Fluroscopy Time	Patient-specific PS > Generic PS > control	$p = 0.022$
		Contrast Volume	NS	
		Number of Roadmaps	NS	
		Time to catheterise the CCA	NS	
		Time to catheterise the ICA	NS	
		Total time that the embolic protection device was deployed in the ICA	Patient specific PS > Generic PS & control	$p < 0.001$
	Subjective Questionnaire	5-point Likert scale	Patient-specific PS > Generic PS > control	$p = 0.017$
Do *et al.*	Timing Score ‡	$\frac{Time\ Taken}{No.\ of\ tablets\ transferred}$	PS > control	$p = 0.0001$
	% difference between warm-up & follow-up	$\frac{Warm\text{-}up - Follow\text{-}up}{Warm\text{-}up} * 100$	PS > control	$p = 0.0001$ & $p = 0.0112$
Moldovanu *et al.*	Global Rating Scale	Total OSATS Score	NS	
		Respect for Tissue	PS > control	$p = 0.034$
		Time & Motion	NS	
		Instrument Handling	NS	
		Depth Perception	NS	
		Bimanual Dexterity	NS	

Authors	Outcome Measure	Details	Results	Significance
		Overall Impression	NS	
Rosser *et al.*	Cobra Rope Task	Time taken to uncoil a string	PS > control	$p < 0.05$
	Triangle Transfer Drill	Time taken & errors made	NS	
	Slam Dunk Drill	Time taken & errors made	NS	
	Interrupted Intracorporeal Sutures	Time taken & errors made	PS > control	$p < 0.05$
	Composite Score	Time taken performing each drill + errors (error = 5s)	NS ¶	
Lee *et al.*	Cognitive Performance	Attention	PS > control	$p < 0.02$
		Distraction / drowsiness	PS > control	$p < 0.001$
		Mental Workload	PS > control	$p < 0.02$
		Index of Cognitive Activity	Unclear	
	Psychomotor Performance	Hand Movement Smoothness	PS > control	$p < 0.03$
		Tool Movement Smoothness	PS > control	$p < 0.05$
		Posture Stability	PS > control	$p < 0.05$
	Global Rating Scale	For Mobilisation of the Colon	PS > control	$p = 0.04$
		For Intracorporeal Suturing & Knot Tying	NS	
Calatayud *et al.*	Global Rating Scale	Total OSATS Score	PS > control	$p = 0.042$
		Respect for Tissue	NS	
		Time & Motion	NS	
		Instrument Handling	NS	
		Knowledge of Instruments	NS	
		Use of Assistants	NS	
		Flow of Operation & Operative Planning	NS	
		Knowledge of Specific Operation	NS	
Lendavay *et al.*	Time	Time to Complete Task	PS > control	$p = 0.001$
	Simulator-generated metrics	Tool Path Length	PS > control	$p = 0.014$
		Economy of Motion	NS	
		Technical Errors	NS	
		Cognitive Errors	NS	
	Global Technical Error	Error Count	PS > control	$p = 0.020$
Kahol *et al.*	Experiment 1	Gesture-level Proficiency	PS > control	$p < 0.005$
		Hand-movement Smoothness	PS > control	$p < 0.005$
		Tool-movement Smoothness	PS > control	$p < 0.005$
		Time Elapsed	PS > control	$p < 0.005$
		Cognitive Errors	PS > control	$p < 0.005$
	Experiment 2	Gesture-level Proficiency	PS > control	$p < 0.001$ & $P < 0.004$
		Hand-movement Smoothness	PS > control	$p < 0.02$ & $p < 0.001$
		Tool-movement Smoothness	Unclear	
		Time Elapsed	NS	
		Cognitive Errors	PS > control	$p < 0.004$ & $p < 0.001$

Authors	Outcome Measure	Details	Results	Significance
Plerhoples *et al.*	Object Positioning Task	Time	NS	
		Path Length	NS	
		Smoothness	NS	
		Hand Dominance	NS	
		Errors	PS > control	$p = 0.01$
	Tissue Manipulating Task	Time	NS	
		Path Length	NS	
		Smoothness	NS	
		Hand Dominance	NS	
		Errors	PS > control	$p = 0.05$
	Total Scores	Time	NS	
		Path Length	NS	
		Smoothness	NS	
		Hand Dominance	NS	
		Errors	PS > control	$p = 0.002$
Weston *et al.*	Global Rating Scale	Depth Perception	NS	
		Bimanual Dexterity	NS	
		Efficiency	NS	
		Tissue Handling	NS	
Chen *et al.*	Global Rating Scale	Reznick subscale (Total Score)	PS > control	$p < 0.001$
		Respect for Tissue	PS > control	$p = 0.005$
		Time & Motion	PS > control	$p = 0.004$
		Instrument Handling	PS > control	$p < 0.001$
		Knowledge of Instruments	PS > control	$p < 0.001$
		Use of Assistants	PS > control	$p = 0.028$
		Knowledge of Specific Procedures	PS > control	$p = 0.001$
		Vassiliou subscale (Total Score)	PS > control	$p < 0.001$
		Depth Perception	PS > control	$p < 0.001$
		Bimanual Dexterity	PS > control	$p < 0.001$
		Efficiency	PS > control	$p < 0.001$
		Tissue Handling	PS > control	$p = 0.004$
		Autonomy	PS > control	$p < 0.001$
		Kundhal subscale (Total Score)	PS > control	$p < 0.001$
		Respect for Tissue	PS > control	$p = 0.028$
		Precision of Operative Technique	PS > control	$p < 0.001$
		Economy of Movement	PS > control	$p < 0.001$
		Confidence of Movements	PS > control	$p = 0.001$
Arora *et al.* ⋆	Global Rating Scale	Total OSATS Score	PS > control	$p < 0.001$
		Respect for Tissue	Not Reported	
		Time & Motion	Not Reported	
		Instrument Handling	Not Reported	
		Knowledge of Instruments	Not Reported	
		Use of Assistants	Not Reported	
		Flow of Operation & Forward Planning	Not Reported	
		Knowledge of Specific Procedure	Not Reported	

Authors	Outcome Measure	Details	Results	Significance
	Mental Imagery Questionnaire	How ready or 'energised' do you feel..?	PS > control	$p < 0.01$
		How confident do you feel..?	PS > control	$p < 0.001$
		How well do you think you can perform... compared with others..?	PS > control	$p < 0.001$
		How helpful is the [PS] activity in preparing you..?	PS > control	$p < 0.001$
		How easily can you 'see' yourself performing..?	PS > control	$p < 0.001$
		How vivid & clear are the images in your mind..?	PS > control	$p < 0.001$
		How easily can you feel yourself performing..?	PS > control	$p < 0.001$
		How easily would you be able to talk someone through the steps..?	PS > control	$p < 0.001$

† Kroft *et al.* report a significant result but only when junior residents were excluded from analysis and when the warm-up was completed in the second session.

‡ Note: if participants failed to complete the task within the allotted time, a different scoring system was used (number of tablets transferred).

¶ Reported as "marginally significant" at $p = 0.07$.

⋆ Results from Session 5 (final session).

Appendix C

Underlying Processes Examined by the Included Studies

Table C.1 Summary of the underlying cognitive processes examined by the included studies

Author	Preoperative Simulation	Learning Processes
Kroft *et al.*	Simple Simulated Task	'Fast'
Willaert *et al.*	Generic Simulation and Specific Simulation	'Fast' & 'Slow'
Do *et al.*	Simple Simulated Task	'Fast'
Moldovanu *et al.*	Simplified Simulated Task	'Fast' & 'Slow'
Rosser *et al.*	Motor System Warm-up (Computer Game)	'Fast'
Lee *et al.*	Simplified Simulated Task	'Fast'
Calatayud *et al.*	Simplified Simulated Task	'Fast'
Lendavay *et al.*	Simplified Simulated Task	'Fast'
Kahol *et al.*	Simplified Simulated Task	'Fast'
Plerhoples *et al.*	Motor System Warm-up (Computer Game)	'Fast'
Weston *et al.*	Simplified Simulated Task and Motor System Warm-up (Computer Game)	'Fast'
Chen *et al.*	Simplified Simulated Task	'Fast'
Arora *et al.*	Mental Practice	'Slow'

Appendix D

Individual Learning Results from Experimentα

Table D.1 Individual Performance during αLearn:K&S.
Results marked with † indicate participants who did not achieve a predicted plateau.

Participant No.	Coefficient of Determination (R^2)	Predicted Plateau	Minimum Achieved
1	0.9425	282078	174000
2	0.9072	5045	235000 †
3	0.9482	78846	125000 †
4	Interrupted	-	-
5	0.9488	94152	164000 †
6	1	1060000	1030000
7	0.994	566365	407000
8	0.9881	193195	131000
9	0.9657	371853	225000
10	0.9975	393726	295000
11	0.9703	171458	127000
12	0.8976	563518	279000
13	0.8193	193442	146000
14	0.9768	288912	147000
15	0.9934	335772	220000
16	0.8688	160486	198000 †
17	0.9803	229611	189000
18	0.8481	464381	223000
19	0.8625	383330	179000
20	0.9824	268816	177000
21	0.9919	180560	176000
22	0.9974	199826	138000
23	0.8931	131865	134000 †
24	0.9862	256263	158000
25	0.7216	137334	128000
26	0.9909	211437	141000
27	0.7255	198583	134000
28	0.9498	165974	104000
29	0.5738	572038	153000
30	0.9981	229843	147000
31	0.9848	230144	149000
32	0.9951	230144	169000
33	0.878	1619000	876000
34	0.5908	139186	107000
35	0.8429	371423	156000
36	0.9928	279865	172000
37	0.9888	209210	127000
38	0.9724	289983	197000

Table D.2 Individual Performance during αLearn:C,K&S.
Results marked with a ‡ indicate a low R^2. Results marked with † indicate participants who did not achieve a predicted plateau.

Participant No.	Coefficient of Determination (R^2)	Predicted Plateau	Minimum Achieved
1	0.5529	191609	129000
2	0.6138	326107	183000
3	0.6788	120555	96000
4	0.01352 ‡	251049	146000
5	0.9963	170957	156000
6	0.8332	-959226	378000
7	0.9455	339257	278000
8	0.6833	185799	146000
9	0.7365	260772	174000
10	0.94	-41570000	265000
11	0.5522	39433	132000 †
12	0.737	203047	175000
13	0.8831	166667	127000
14	0.6742	239329	170000
15	0.6742	239329	170000 †
16	0.6964	230751	146000
17	0.8723	-85570000	169000
18	0.9752	-333300000	206000
19	0.8935	200059	140000
20	Interrupted	-	-
21	0.9999	151761	177000 †
22	0.9999	151761	177000 †
23	0.7194	-36408	130000
24	0.1912 ‡	219864	149000
25	0.5507	256263	98500
26	0.522	123630	133000 †
27	0.9426	126934	111000
28	0.2137 ‡	60497	110000
29	0.823	233299	147000
30	0.7001	-82430000	149000
31	0.7034	-151400000	134000
32	0.5382	65324	141000 †
33	0.8175	675711	518000
34	0.8144	131611	112000
35	0.8543	-59988	117000
36	0.6078	238871	131000
37	0.9062	125989	118000
38	0.9749	186602	160000

Table D.3 Individual Performance during αLearn:C.
Results marked with a ‡ indicate a low R^2. Results marked with † indicate participants who did not achieve a predicted plateau.

Participant No.	Coefficient of Determination (R^2)	Predicted Plateau	Minimum Achieved
1	0.1825 ‡	8444	3850
2	0.9682	6560	4260
3	0.2041 ‡	1344	1570 †
4	0.8653	5242	2870
5	0.6716	2736	1730
6	0.6246	14682	10600
7	0.9726	3531	2260
8	Interrupted	-	-
9	0.3935	-3919	1930
10	0.4207	-96828	1310
11	0.9319	3067	1050
12	0.7793	512.1	1770 †
13	0.8409	1212	1670
14	0.01191 ‡	-167500000	4630
15	0.7767	3965	2970
16	0.9805	7622	2970
17	0.8791	2447	1630
18	Interrupted	-	-
19	0.6339	1912	2480 †
20	0.5644	-5660000	4270
21	Interrupted	-	-
22	Interrupted	-	-
23	0.6703	1718	1260
24	0.8146	4161	2430
25	0.6519	2557	1440
26	0.6174	1785	3590 †
27	0.7016	-1570000	1960
28	0.04231 ‡	3378	920
29	0.7628	3417	3060
30	0.5884	-6661	1670
31	0.5989	2313	1580
32	0.6985	4153	1760
33	Interrupted	-	-
34	0.2241 ‡	2278	1350
35	0.389	4061	1960
36	Interrupted	-	-
37	0.8252	2564	1200
38	Interrupted	-	-

Appendix E

Instructions to Participants (Experimentα)

E.1 Learning-Phase Instructions

E.1.1 Learning Sequence Information

Please enter your participant code. Once you press return, you will see 16 numbered, coloured squares in a circle. The aim of this task is to move a pen-like stylus to each square, in order. You start the task by clicking on the screen. Once you do this, the numbers will disappear. You need to remember the order of squares and move to each in order. Once you reach the correct square, it will disappear.

We would like you to be both quick and accurate. During this part of the experiment, we aim to teach you the colour of the squares and so you can repeat the task as many times as you'd like. However, **each time you repeat the task, the position of the squares in the circle will change**. The numbers will always stay the same (i.e. Square 1 will always be blue).

E.1.2 Learning Motoric Information

Please enter your participant code. Once you press return, you will see 16 numbered squares distributed across a grid. The aim of this task to to move the laparoscopic instrument to each square, in order. You start the task by clicking on the screen. Once you do this, the numbers will disappear. You need to remember the order of squares and move to each in order. Once you reach the correct square, it will disappear.

We would like you to be both quick and accurate. During this part of the experiment, we aim to teach you the location of the squares and so you can repeat the task as many times as you'd like.

E.1.3 Learning Combined Information

Please enter your participant code. Once you press return, you will see 16 numbered, coloured squares distributed across a grid. The aim of this task to to move the laparoscopic instrument to each square, in order. You start the task by clicking on the screen. Once you do this, the numbers will disappear. You need to remember the order of squares and move to each in order. Once you reach the correct square, it will disappear.

We would like you to be both quick and accurate. During this part of the experiment, we aim to teach you the sequence of the squares and so you can repeat the task as many times as you'd like.

E.2 Assessment of Preoperative Simulation Instructions

E.2.1 Warm-up A (αPS:K&S)

Please enter your participant number. Once you press return, you will see 16 numbered squares distributed across a grid. The aim of this task to to move the laparoscopic instrument to each square, in order. You start the task by clicking on the screen. Once you do this, the numbers will disappear. You need to remember the order of squares and move to each in order. Once you reach the correct square, it will disappear.

We would like you to be both quick and accurate. You have **three minutes** to perform this warm-up task, during which time you can repeat the task as many times as you'd like.

After three minutes, you will see another 16 coloured squares distributed across a grid. This is the same task that you learnt during the earlier part of this experiment. The aim of this task to to move the laparoscopic instrument to each square, in order. You start the task by clicking on the screen. You need to remember the order of squares from before and move to each in order. Once you reach the correct square, it will disappear.

We would like you to be both quick and accurate. **This is an assessed task; you will only have one attempt**.

E.2.2 Warm-up B (αPS:C)

Please enter your participant number. Once you press return, you will see 16 numbered, coloured squares in a circle. The aim of this task to to move the mouse to each square, in order. You start the task by clicking on the screen. Once you do this, the numbers will disappear. You need to remember the order of squares and move to each in order. Once you reach the correct square, it will disappear.

We would like you to be both quick and accurate. You have **three minutes** to perform

this warm-up task, during which time you can repeat the task as many times as you'd like. Please remember that **each time you repeat the task, the position of the squares in the circle will change**.

After three minutes, you will see another 16 coloured squares distributed across a grid. This is the same task that you learnt during the earlier part of this experiment. The aim of this task to to move the laparoscopic instrument to each square, in order. You start the task by clicking on the screen. You need to remember the order of squares from before and move to each in order. Once you reach the correct square, it will disappear.

We would like you to be both quick and accurate. **This is an assessed task; you will only have one attempt**.

E.2.3 Warm-up C (αPS:O)

For this trial, you will see a video of someone performing the task you learnt during the earlier part of this experiment. You have **three minutes to review the video**.

After three minutes, you will see another 16 coloured squares distributed across a grid. This is the same task that you learnt during the earlier part of this experiment and the same as shown in the video. The aim of this task to to move the laparoscopic instrument to each square, in order. You start the task by clicking on the screen. You need to remember the order of squares from before and move to each in order. Once you reach the correct square, it will disappear.

We would like you to be both quick and accurate. **This is an assessed task; you will only have one attempt**.

E.2.4 Warm-up D (αPS:K,S&C)

Please enter your participant number. Once you press return, you will see 16 numbered, coloured squares distributed across a grid. This is the same task that you learnt during the earlier part of this experiment. The aim of this task to to move the laparoscopic instrument to each square, in order. You start the task by clicking on the screen. You need to remember the order of squares from before and move to each in order. Once you reach the correct square, it will disappear.

We would like you to be both quick and accurate. You have **three minutes** to perform this warm-up task, during which time you can repeat the task as many times as you'd like.

After three minutes, you will see the same 16 coloured squares distributed across a grid. The aim of this task to to move the laparoscopic instrument to each square, in order. You start the task by clicking on the screen. You need to remember the order of squares from before and move to each in order. Once you reach the correct square, it will disappear.

We would like you to be both quick and accurate. **This is an assessed task; you will only have one attempt**.

E.2.5 Warm-up E (Control)

For this trial, you get to play the game Tetris. You have **three minutes** to play the game using the computer keyboard in front of you.

After three minutes, you will see another 16 coloured squares distributed across a grid. This is the same task that you learnt during the earlier part of this experiment. The aim of this task to to move the laparoscopic instrument to each square, in order. You start the task by clicking on the screen. You need to remember the order of squares from before and move to each in order. Once you reach the correct square, it will disappear.

We would like you to be both quick and accurate. **This is an assessed task; you will only have one attempt**.

Appendix F

Individual Learning Results from Experimentβ

Table F.1 Individual Performance during βLearn:K.
Results marked with † indicate participants who did not achieve a predicted plateau.

Participant No.	Coefficient of Determination (R^2)	Predicted Plateau	Minimum Achieved
1	0.831	3413	2250
2	0.9184	2798	1720
3	0.8562	2471	2070
4	0.6629	2304	1570
5	Interrupted	-	-
6	Interrupted	-	-
7	0.958	2751	1850
8	0.9712	1940	1600
9	0.8908	4268	2620
10	0.6488	2011	1990
11	Interrupted	-	-
12	0.9451	1740	1460
13	0.7002	1930	1320
14	0.621	1697	1780 †
15	0.9595	5533	3440
16	0.8856	2034	1960
17	0.8162	2109	1640
18	0.6021	-76785	2160
19	0.9255	3967	3540
20	0.4432	2887	1886
21	0.8739	3624	2950
22	0.6428	-138681	2810
23	0.9285	2955	2050
24	Interrupted	-	-
25	0.5215	3491	2740
26	0.6747	2178	1600
27	0.7132	4048	2870
28	0.7255	3020	2840
29	0.9983	5725	3590
30	0.7797	2337	2040
31	0.9279	1963	1630
32	0.728	1146	1750 †
33	0.5658	3324	2690
34	0.9335	2725	2210
35	0.9598	734.7	3520 †
36	Interrupted	-	-

Table F.2 Individual Performance during βLearn:C&K.
Results marked with a ‡ indicate a low R^2. Results marked with † indicate participants who did not achieve a predicted plateau.

Participant No.	Coefficient of Determination (R^2)	Predicted Plateau	Minimum Achieved
1	0.8566	3290	1880
2	0.7649	2767	1800
3	0.8155	2194	1760
4	0.9094	2732	1550
5	0.6472	11294	2920
6	Interrupted	-	-
7	0.4281‡	713	1470 †
8	0.5085	1509	1130
9	Interrupted	-	-
10	0.7006	3012	1790
11	Interrupted	-	-
12	0.2137 ‡	1811	1400
13	0.303 ‡	1262	1130
14	Interrupted	-	-
15	0.6528	-3498000	3280
16	0.7923	1184	1940 †
17	Interrupted	-	-
18	Interrupted	-	-
19	0.6405	1740	2710 †
20	Interrupted	-	-
21	Interrupted	-	-
22	Interrupted	-	-
23	Interrupted	-	-
24	0.9217	-20851	3670
25	Interrupted	-	-
26	0.243 ‡	2149	1380
27	0.9734	5530	6360 †
28	Interrupted	-	-
29	Interrupted	-	-
30	Interrupted	-	-
31	0.04266 ‡	1872	1440
32	0.197 ‡	1832	1420
33	Interrupted	-	-
34	0.6432	2518	1840
35	Interrupted	-	-
36	0.7974	7666	4070

Table F.3 Individual Performance during βLearn:C.
Results marked with a ‡ indicate a low R^2. Results marked with † indicate participants who did not achieve a predicted plateau.

Participant No.	Coefficient of Determination (R^2)	Predicted Plateau	Minimum Achieved
1	0.4118 ‡	-1239000	1670
2	0.6144	1141	1520 †
3	0.9404	1548	1080
4	Interrupted	-	-
5	0.6447	2671	1620
6	0.8044	997.9	525
7	Interrupted	-	-
8	0.7878	1365	688
9	0.05696 ‡	3043	1050
10	0.5167	-653388	2080
11	0.8477	5710	1680
12	0.9199	909.1	521
13	0.7637	885.8	601
14	0.9047	1135	766
15	Interrupted	-	-
16	0.7073	-2386000	878
17	0.754	1884	1150
18	0.6319	-556.6	1210
19	Interrupted	-	-
20	0.8652	1246	733
21	0.4035 ‡	432.1	1090 †
22	Interrupted	-	-
23	0.1639 ‡	2391	1160
24	0.3982	1488	686
25	0.5357	-93864	851
26	0.7933	1397	666
27	Interrupted	-	-
28	0.9613	1438	1260
29	0.8438	1132	922
30	Interrupted	-	-
31	Interrupted	-	-
32	0.5359	1032	637
33	0.5488	1351	816
34	Interrupted	-	-
35	0.01584 ‡	2993	1220
36	0.5418	533.7	944 †

Appendix G

Individual Learning Results from Experimentγ

Table G.1 Individual Performance during γLearn:C.
Results marked with a ‡ indicate a low R^2. Results marked with † indicate participants who did not achieve a predicted plateau.

Participant No.	Coefficient of Determination (R^2)	Predicted Plateau	Minimum Achieved
1	0.05693 ‡	2.256	4.32 †
2	Interrupted	-	-
3	Interrupted	-	-
4	Interrupted	-	-
5	0.1822 ‡	9.123	4.3
6	Interrupted	-	-
7	Interrupted	-	-
8	0.8135	8.019	5.31
9	0.7576	8.703	4.4
10	0.2769 ‡	3.083	3.36 †
11	0.5402	5.166	3.76
12	0.5454	7.4	5.02
13	0.2692 ‡	3.878	6.24 †
14	0.2472 ‡	4.874	3.31
15	0.8347	4.837	3.47
16	0.6361	-3896	7.03 †
17	0.5304	5.968	3.15
18	0.3725 ‡	6.765	4.51
19	0.8738	4.171	3.58
20	0.1894 ‡	2.047	3.49 †
21	Interrupted	-	-
22	0.9498	7.497	4.76
23	0.371 ‡	6.708	3.02
24	Interrupted	-	-
25	0.4469 ‡	7.144	3.04
26	0.7147	9.56	4.85
27	0.44 ‡	4.867	3.56
28	0.5904	8.534	3.88
29	0.844	5.185	2.93
30	0.7581	9.044	4.42
31	0.2392 ‡	8.099	5.41
32	0.9577	6.714	2.88
33	Interrupted	-	-
34	0.07585 ‡	6.519	3.03
35	0.6291	4.108	2.62
36	0.5218	3.26	2
37	0.7337	6.216	3.17
38	0.08294 ‡	7.472	3.15
39	0.06069 ‡	5.351	3.21
40	0.2971 ‡	4.17	2.43
41	Interrupted	-	-
42	0.3701 ‡	7.374	4.28
43	0.7652	6.739	4.74

Table G.2 Individual Performance during γLearn:C,K&S.
Results marked with a ‡ indicate a low R^2. Results marked with † indicate participants who did not achieve a predicted plateau.

Participant No.	Coefficient of Determination (R^2)	Predicted Plateau	Minimum Achieved
1	0.5877	33447	28000
2	0.679	43592	33400
3	0.5838	34898	24100
4	0.002902 ‡	913400000	52100
5	0.5838	34898	24100
6	0.4725 ‡	49558	29500
7	0.9521	90606	51100
8	0.9968	65542	53700
9	0.809	62151	42800
10	0.8547	29375	29200
11	0.9111	44604	32500
12	0.927	41258	29100
13	0.7066	52419	36900
14	0.4992 ‡	48927	33900
15	0.8608	46783	28200
16	0.058 ‡	68080	41300
17	0.1209 ‡	56837	28000
18	0.9666	41746	28300
19	Interrupted	-	-
20	0.998	39746	27200
21	0.4717 ‡	42033	27400
22	0.8008	41438	34200
23	0.8741	35092	25300
24	0.5087	32983	28500
25	0.5395	13257	18700 †
26	0.8131	89193	50000
27	0.8444	56097	36800
28	0.8441	46795	35100
29	0.6342	27956	21300
30	Interrupted	-	-
31	0.7316	58698	40100
32	0.9128	96731	54600
33	0.6537	34349	29500
34	Interrupted	-	-
35	0.8808	34936	21400
36	0.6141	33768	23800
37	0.8156	40357	29300
38	Interrupted	-	-
39	0.2936 ‡	50578	39100
40	0.459 ‡	38309	24800
41	Interrupted	-	-
42	0.7389	62226	47400
43	0.911	46884	35100

Table G.3 Individual Performance during γLearn:C,K&FS.
Results marked with a ‡ indicate a low R^2. Results marked with † indicate participants who did not achieve a predicted plateau.

Participant No.	Coefficient of Determination (R^2)	Predicted Plateau	Minimum Achieved
1	0.001957 ‡	14.18	8.62
2	0.285 ‡	7.619	10 †
3	0.08494 ‡	10.56	8.24
4	0.4138 ‡	15.02	12
5	Interrupted	-	-
6	Interrupted	-	-
7	0.3615 ‡	21.41	16.4
8	Interrupted	-	-
9	0.1021 ‡	62151	42800
10	0.3468 ‡	-77.51	8.17 †
11	0.1739 ‡	6.554	8.96 †
12	Interrupted	-	-
13	Interrupted	-	-
14	0.1341 ‡	13.47	9.84
15	Interrupted	-	-
16	Interrupted	-	-
17	0.08419 ‡	13.78	8.84
18	Interrupted	-	-
19	0.1719 ‡	-0.01834	9.14 †
20	0.1345 ‡	9.512	7.95
21	0.9425	11.33	9.14
22	0.4512 ‡	12.38	11
23	0.5666	11.34	6.94
24	Interrupted	-	-
25	Interrupted	-	-
26	0.3416 ‡	-129.5	14.4 †
27	0.2179 ‡	13.67	11.7
28	0.6668	13.66	11.1
29	0.3879 ‡	8.902	7.19
30	Interrupted	-	-
31	Interrupted	-	-
32	Interrupted	-	-
33	0.2722 ‡	8.166	9.58 †
34	0.4615 ‡	12.44	8.84
35	Interrupted	-	-
36	0.1518 ‡	10.61	8.79
37	Interrupted	-	-
38	Interrupted	-	-
39	0.03023 ‡	15.59	10.5
40	0.08697 ‡	4.289	7.4 †
41	0.9543	9.829	8.03
42	Interrupted	-	-
43	0.5833	12.03	8.44

Appendix H

Spire Research Ethics Committee Approval

Spire Healthcare

Spire Healthcare Ltd
Head Office
3 Dorset Rise
London
EC4Y 8EN

13th February 2015
Mr Pike
Spire Leeds Hospital
Jackson Avenue
Leeds
LS8 1NT

Dear Mr Pike

Project Title: Retrospective analysis of real-world instances of a preoperative warm-up
Spire REC ID: SRECSE01

Confirmation of ethical opinion
On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above service evaluation on the basis described in the application form and supporting documentation, as revised, subject to the conditions specified below. Please see the enclosed "Communication of Spire Decision to Applicant" for further details.

Conditions of the favourable opinion
The favourable opinion is subject to the following conditions being met prior to the start of the audit. Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned, therefore please inform the local MAC of the project activity.

Statement of compliance
The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committee and complies fully with the Clinical Policy 06 Research.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Reporting requirements
The reporting requirements for studies with a favourable opinion include:

Notifying substantial amendments

- Adding changes substantial changes to the project, adding new sites and investigators (immediate notification)
- Progress and safety reports (annually, refer to enclosed documents and Spire Policy HOp 03 Health and Safety Policy)
- Notifying the end of the study (enclosed document)

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Supporting Document Included:	*Resources / Guidelines for document content:*	YES	N/A
Protocol	*Guidelines: ICH GCP E6, Section 6. CLINICAL TRIAL PROTOCOL http://ichgcp.net/6-clinical-trial-protocol-and-protocol-amendments*	X	
Consent Form(s) & Patient Information Sheet(s)	*Guidance and Templates: http://www.hra-decisiontools.org.uk/consent/*		X
Indemnity certificate of lead consultant (e.g. GMC)	*Spire Clinical Policy 05*	X	
If Sponsor, Sponsor indemnity	*Spire Clinical Policy 05*		X
Research Contract / SLA	*If the sponsor has not produced a contract or SLA, Spire can advise spireethics@spirehealthcare.com See Spire clinical intranet for legal templates for (research) service provision (NHS & Non NHS Spot Purchasing Agreements and Guidelines) http://intranet.spirehealthcare.net/Left-Menu/Legal/*		X
Lead Investigator C.V. (incl. copy of GCP training certificate)		X	
Copy of existing NHS LREC approval (if applicable)			X

With the Committee's best wishes for the success of this project.

Yours sincerely

Catherine Sinfield

SPIRE REC Coordinator

Email: spireethics@spirehealthcare.com

Enclosures:

"ETHICS CHECKLIST DECISION"

"ANNUAL PROGRESS REPORT TO SPIRE RESEARCH ETHICS COMMITTEE"

"DECLARATION OF THE END OF A STUDY"

Copy to: Wendy Bartle, Spire Leeds Hospital

References

[1] Weiser TG, Regenbogen SE, Thompson KD, Haynes AB, Lipsitz SR, Berry WR, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. Lancet. 2008;372(9633):139–144.

[2] Population Reference Bureau. 2006 world population data sheet; 2006. `http://www.prb.org/pdf06/06WorldDataSheet.pdf`.

[3] World Health Organization. WHO Guidelines for Safe Surgery 2009: Safe Surgery Saves Lives. Geneva; 2009.

[4] Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. Surgery. 1999;126(1):66–75.

[5] Kable AK, Gibberd RW, Spigelman AD. Adverse events in surgical patients in Australia. Int J Qual Health Care. 2002;14(4):269–276.

[6] Bickler SW, Sanno-Duanda B. Epidemiology of paediatric surgical admissions to a government referral hospital in the Gambia. Bull World Health Organ. 2000;78(11):1330–1336.

[7] Yii MK, Ng KJ. Risk-adjusted surgical audit with the POSSUM scoring system in a developing country. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. Br J Surg. 2002;89(1):110–113.

[8] McConkey SJ. Case series of acute abdominal surgery in rural Sierra Leone. World J Surg. 2002;26(4):509–513.

[9] Ronsmans C, Graham WJ. Maternal mortality: who, when, where, and why. Lancet. 2006;368(9542):1189–1200.

[10] de Vries EN, Prins HA, Crolla RMPH, den Outer AJ, van Andel G, van Helden SH, et al. Effect of a comprehensive surgical safety system on patient outcomes. N Engl J Med. 2010;363(20):1928–1937.

[11] The Royal College of Surgeons of England. Good Surgical Practice. London; 2014.

[12] Baldwin PJ, Paisley AM, Brown SP. Consultant surgeons' opinion of the skills required of basic surgical trainees. Br J Surg. 1999;86(8):1078–1082.

[13] Yule S, Flin R, Paterson-Brown S, Maran N, Rowley D. Development of a rating system for surgeons' non-technical skills. Med Educ. 2006;40(11):1098–1104.

[14] Crossley J, Marriott J, Purdie H, Beard JD. Prospective observational study to evaluate NOTSS (Non-Technical Skills for Surgeons) for assessing trainees' non-technical performance in the operating theatre. Br J Surg. 2011;98(7):1010–1020.

[15] Clark D, Ivry RB. Multiple systems for motor skill learning. Wiley Interdiscip Rev Cogn Sci. 2010;1(4):461–467.

[16] Poldrack RA, Clark J, Paré-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, et al. Interactive memory systems in the human brain. Nature. 2001;414(6863):546–550.

[17] Beilock SL, Carr TH, MacMahon C, Starkes JL. When paying attention becomes counterproductive: impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. J Exp Psychol Appl. 2002;8(1):6–16.

[18] Wulf G, Weigelt C. Instructions about physical principles in learning a complex motor skill: to tell or not to tell.... Res Q Exerc Sport. 1997;68(4):362–367.

[19] Fitts PM. The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol. 1954;47(6):381–391.

[20] Hick WE. On the rate of gain of information. Q J Exp Psychol. 1952;4(1):11–26.

[21] Hyman R. Stimulus information as a determinant of reaction time. J Exp Psychol. 1953;45(3):188–196.

[22] Seibel R. Discrimination reaction time for a 1,023-alternative task. J Exp Psychol. 1963;66:215–226.

[23] Newell A, Rosenbloom PS. Mechanisms of skill acquisition and the law of practice. In: Anderson J, editor. Cognitive Skills and Their Acquisition. Hillsdale, NJ; 1981. .

[24] Logan GD. Shapes of reaction-time distributions and shapes of learning curves: a test of the instance theory of automaticity. J Exp Psychol Learn Mem Cogn. 1992;18(5):883–914.

[25] Heathcote A, Brown S, Mewhort DJ. The power law repealed: the case for an exponential law of practice. Psychon Bull Rev. 2000;7(2):185–207.

[26] Rosenbloom PS, Newell A. Learning by chunking: A production-system model of practice. In: Klahr D, Langley P, Neches R, editors. Self-modifying production system models of learning and development. Cambridge, MA: MIT Press; 1987. .

[27] Ramkumar P, Acuna DE, Berniker M, Grafton ST, Turner RS, Körding KP. Chunking as the result of an efficiency computation trade-off. Nat Commun. 2016;7:1–11.

[28] MacKay DG. The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. Psychol Rev. 1982;89(5):483–506.

[29] Verwey WB. Concatenating familiar movement sequences: the versatile cognitive processor. Acta Psychol. 2001;106:69–95.

[30] Verwey WB, Wright DL. Effector-independent and effector-dependent learning in the discrete sequence production task. Psychol Res. 2004;68(1):64–70.

[31] Abrahamse EL, Ruitenberg MFL, de Kleine E, Verwey WB. Control of automated behavior: insights from the discrete sequence production task. Front Hum Neurosci. 2013;7:82.

[32] Verwey WB. Evidence for a multistage model of practice in a sequential movement task. J Exp Psychol Hum Percept Perform. 1999;25(6):1693–1708.

[33] Tattersall AJ, Broadbent DE. Output buffer storage and the modality of recall. Q J Exp Psychol [A]. 1991;43(1):1–18.

[34] Smyth MM, Pendleton LR. Working memory for movements. Q J Exp Psychol [A]. 1989;41(2):235–250.

[35] Khan M, Snelling A, Tiernan E. The need for technical skills assessment in surgery. Int J Surg. 2005;3(1):83–86.

[36] Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev. 1993;100(3):363–406.

[37] Diedrichsen J, Kornysheva K. Motor skill learning between selection and execution. Trends Cogn Sci. 2015;19(4):227–233.

[38] Shadmehr R, Wise S. The computational neurobiology of reaching and pointing: A foundation for motor learning. MIT Press; 2005.

[39] Flash T, Hogan N. The coordination of arm movements: an experimentally confirmed mathematical model. J Neurosci. 1985;5(7):1688–1703.

[40] Todorov E, Jordan MI. Smoothness maximization along a predefined path accurately predicts the speed profiles of complex arm movements. J Neurophysiol. 1998;80(2):696–714.

[41] Stewart J. Calculus: Early transcendentals. 6th ed. Belmont, CA: Brooks/Cole; 2007.

[42] Nissen MJ, Bullemer P. Attentional requirements of learning: Evidence from performance measures. Cogn Psychol. 1987;19(1):1–32.

[43] Cohen A, Ivry RI, Keele SW. Attention and structure in sequence learning. J Exp Psychol Learn Mem Cogn. 1990;16(1):17–30.

[44] Keele SW, Jennings P, Jones S, Caulton D, Cohen A. On the modularity of sequence representation. J Mot Behav. 1995;27(1):17–30.

[45] Karni A, Meyer G, Jezzard P, Adams MM, Turner R, Ungerleider LG. Functional MRI evidence for adult motor cortex plasticity during motor skill learning. Nature. 1995;377(6545):155–158.

[46] Pike TW, Pathak S, Mushtaq F, Wilkie RM, Mon-Williams M, Lodge JPA. A systematic examination of preoperative surgery warm-up routines. Surg Endosc. 2016;92(4):102–2214.

[47] Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ. A Systematic Review of Skills Transfer After Surgical Simulation Training. Ann Surg. 2008;248(2):166–179.

[48] Dawe SR, Pena GN, Windsor JA, Broeders JAJL, Cregan PC, Hewett PJ, et al. Systematic review of skills transfer after surgical simulation-based training. Br J Surg. 2014;101(9):1063–1076.

[49] Herrell SD, Smith JR JA. Robotic-assisted laparoscopic prostatectomy: What is the learning curve? Urology. 2005;66(5):105–107.

[50] Tekkis PP, Senagore AJ, Delaney CP, Fazio VW. Evaluation of the learning curve in laparoscopic colorectal surgery: comparison of right-sided and left-sided resections. Ann Surg. 2005;242(1):83.

[51] Tekkis PP, Fazio VW, Lavery IC, Remzi FH, Senagore AJ, Wu JS, et al. Evaluation of the learning curve in ileal pouch-anal anastomosis surgery. Ann Surg. 2005;241(2):262–268.

[52] Lodge JPA, Dasgupta D, Prasad KR, Attia M, Toogood GJ, Davies M, et al. Emergency subtotal hepatectomy: a new concept for acetaminophen-induced acute liver failure: temporary hepatic support by auxiliary orthotopic liver transplantation enables long-term success. Ann Surg. 2008;247(2):238–249.

[53] Sakai K, Kitaguchi K, Hikosaka O. Chunking during human visuomotor sequence learning. Exp Brain Res. 2003;152(2):229–242.

[54] Smith MA, Ghazizadeh A, Shadmehr R. Interacting adaptive processes with different timescales underlie short-term motor learning. PLoS Biol. 2006;4(6):e179–9.

[55] Mazzoni P, Krakauer JW. An implicit plan overrides an explicit strategy during visuomotor adaptation. J Neurosci. 2006;26(14):3642–3645.

[56] Akhtar M. Training to be a surgeon takes too long in the UK. BMJ Careers. 2015;Available from: `http://careers.bmj.com/careers/advice/Training_to_be_a_surgeon_takes_too_long_in_the_UK`.

[57] Karni A, Tanne D, Rubenstein BS, Askenasy JJ, Sagi D. Dependence on REM sleep of overnight improvement of a perceptual skill. Science. 1994;265(5172):679–682.

[58] Benson K, Feinberg I. The beneficial effect of sleep in an extended Jenkins and Dallenbach paradigm. Psychophysiology. 1977;14(4):375–384.

[59] Walker MP, Brakefield T, Morgan A, Hobson JA, Stickgold R. Practice with sleep makes perfect: sleep-dependent motor skill learning. Neuron. 2002;35(1):205–211.

[60] Stickgold R, Hobson JA, Fosse R, Fosse M. Sleep, learning, and dreams: off-line memory reprocessing. Science. 2001;294(5544):1052–1057.

[61] Krakauer JW, Ghez C, Ghilardi MF. Adaptation to visuomotor transformations: consolidation, interference, and forgetting. J Neurosci. 2005;25(2):473–478.

[62] Chalmers CR, Joshi S, Bentley PG, Boyle NH. The lost generation: impact of the 56-hour EWTD on current surgical training. Ann R Coll Surg Engl (Suppl). 2010;92(3):102–106.

[63] Healy NA, Chang KH, Conneely JB, Malone C, Kerin MJ. Impact of supervised training on the acquisition of simulated laparoscopic skills. Ann R Coll Surg Engl (Suppl). 2013;95:1–6.

[64] Ahmed N, Devitt KS, Keshet I, Spicer J, Imrie K, Feldman L, et al. A systematic review of the effects of resident duty hour restrictions in surgery: impact on resident wellness, training, and patient outcomes. Ann Surg. 2014;259(6):1041–1053.

[65] Fradkin AJ, Zazryn TR, Smoliga JM. Effects of warming-up on physical performance: A systematic review with meta-analysis. The Journal of Strength & Conditioning Research. 2010;24(1):140–148.

[66] Feltz DL, Landers DM. The effects of mental practice on motor skill learning and performance: A meta-analysis. J Sport Psychol. 1983;5(1):25–57.

[67] Abdalla G, Moran-Atkin E, Chen G, Schweitzer MA, Magnuson TH, Steele KE. The effect of warm-up on surgical performance: a systematic review. Surg Endosc. 2015;29(6):1259–1269.

[68] Moran-Atkin E, Abdalla G, Chen G, Magnuson TH, Lidor AO, Schweitzer MA, et al. Preoperative warm-up the key to improved resident technique: a randomized study. Surg Endosc. 2014;29(5):1057–1063.

[69] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care Interventions: explanation and elaboration. Ann Intern Med. 2009;151(4):W–65–W–94.

[70] Higgins J, Altman D, Sterne J. Chapter 8: Assessing risk of bias in included studies. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. The Cochrane Collaboration; 2011. .

[71] Mills EJ, Chan AW, Wu P, Vail A, Guyatt GH, Altman DG. Design, analysis, and presentation of crossover trials. Trials. 2009;10(1):27.

[72] Weston MK, Stephens JH, Schafer A, Hewett PJ. Warm-up before laparoscopic surgery is not essential. ANZ J Surg. 2014;84(3):143–147.

[73] Chen CCG, Green IC, Colbert-Getz JM, Steele K, Chou B, Lawson SM, et al. Warm-up on a simulator improves residents' performance in laparoscopic surgery: a randomized trial. Int Urogynecol J. 2013;24(10):1615–1622.

[74] Plerhoples TA, Zak Y, Hernandez-Boussard T, Lau J. Another use of the mobile device: warm-up for laparoscopic surgery. J Surg Res. 2011;170(2):185–188.

[75] Lendvay TS, Brand TC, White L, Kowalewski T, Jonnadula S, Mercer LD, et al. Virtual reality robotic surgery warm-up improves task performance in a dry laboratory environment: a prospective randomized controlled study. JACS. 2013;216(6):1181–1192.

[76] Kroft J, Ordon M, Arthur R, Pittini R. Does surgical "warming up" improve laparoscopic simulator performance? Simul Healthc. 2012;7(6):339–342.

[77] Willaert WIM, Aggarwal R, Daruwalla F, Van Herzeele I, Darzi AW, Vermassen FE, et al. Simulated procedure rehearsal is more effective than a preoperative generic warm-up for endovascular procedures. Ann Surg. 2012;255(6):1184–1189.

[78] Lee JY, Mucksavage P, Kerbl DC, Osann KE, Winfield HN, Kahol K, et al. Laparoscopic warm-up exercises improve performance of senior-level trainees during laparoscopic renal surgery. J Endourol. 2012;26(5):545–550.

[79] Calatayud D, Arora S, Aggarwal R, Kruglikova I, Schulze S, Funch-Jensen P, et al. Warm-up in a virtual reality environment improves performance in the operating room. Ann Surg. 2010;251(6):1181–1185.

[80] Moldovanu R, Târcoveanu E, Dimofte G, Lupaşcu C, Bradea C. Preoperative warm-up using a virtual reality simulator. JSLS. 2011;15(4):533–538.

[81] Rosser JC, Gentile DA, Hanigan K, Danner OK. The effect of video game "warm-up" on performance of laparoscopic surgery tasks. JSLS. 2012;16(1):3–9.

[82] Do AT, Cabbad MF, Kerr A, Serur E, Robertazzi RR, Stankovic MR. A warm-up laparoscopic exercise improves the subsequent laparoscopic performance of Ob-Gyn residents: a low-cost laparoscopic trainer. JSLS. 2006;10(3):297–301.

[83] Kahol K, Satava RM, Ferrara J, Smith ML. Effect of short-term pretrial practice on surgical proficiency in simulated environments: a randomized trial of the "preoperative warm-up" effect. JACS. 2009;208(2):255–268.

[84] Arora S, Aggarwal R, Sirimanna P, Moran A, Grantcharov T, Kneebone R, et al. Mental practice enhances surgical technical skills: A randomized controlled study. Ann Surg. 2011;253(2):265–270.

[85] Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL. Development of a model for training and evaluation of laparoscopic skills. Am J Surg. 1998;175(6):482–487.

[86] Van Herzeele I, Aggarwal R, Malik I, Gaines P. Validation of video-based skill assessment in carotid artery stenting. Eur J Vasc Endovasc Surg. 2009;38(1):1–9.

[87] Martin RCG, Scoggins CR, McMasters KM. Laparoscopic hepatic lobectomy: advantages of a minimally invasive approach. JACS. 2010;210(5):627–34– 634–6.

[88] Kundhal PS, Grantcharov TP. Psychomotor performance measured in a virtual environment correlates with technical skills in the operating room. Surg Endosc. 2009;23(3):645–649.

[89] Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg. 2005;190(1):107–113.

[90] Blencowe NS, McNair AGK, Davis CR, Brookes ST, Blazeby JM. Standards of outcome reporting in surgical oncology: a case study in esophageal cancer. Ann Surg Oncol. 2012;19(13):4012–4018.

[91] Reznick RK. Teaching and testing technical skills. Am J Surg. 1993;165(3):358–361.

[92] Whistance RN, Forsythe RO, McNair AGK, Brookes ST, Avery KNL, Pullyblank AM, et al. A systematic review of outcome reporting in colorectal cancer surgery. Colorectal Dis. 2013;15(10):e548–60.

[93] Macefield RC, Boulind CE, Blazeby JM. Selecting and measuring optimal outcomes for randomised controlled trials in surgery. Langenbecks Arch Surg. 2014;399(3):263–272.

[94] Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. PLoS ONE. 2014;9(6):e99111.

[95] Al-Saud LM, Mushtaq F, Allsop MJ, Culmer PC, Mirghani I, Yates E, et al. Feedback and motor skill acquisition using a haptic dental simulator. Eur J Dent Educ. 2017;21(4):1–8.

[96] Allen RJ, Waterman AH. How does enactment affect the ability to follow instructions in working memory? Mem Cognit. 2015;43(3):555–561.

[97] Fisher RA, Yates F. Statistical tables for biological, agricultural and medical research. London, UK: Oliver and Boyed; 1974.

[98] Herman JP, Bogadhi AR, Krauzlis RJ. Effects of spatial cues on color-change detection in humans. J Vis. 2015;15(6):3.

[99] Carrasco M. Visual attention: the past 25 years. Vision Res. 2011;51(13):1484–1525.

[100] Boynton GM. A framework for describing the effects of attention on visual responses. Vision Res. 2009;49(10):1129–1143.

[101] Martinez-Trujillo JC, Treue S. Feature-based attention increases the selectivity of population responses in primate visual cortex. Curr Biol. 2004;14(9):744–751.

[102] Treue S, Martínez Trujillo JC. Feature-based attention influences motion processing gain in macaque visual cortex. Nature. 1999;399(6736):575–579.

[103] Haenny PE, Schiller PH. State dependent activity in monkey visual cortex. Exp Brain Res. 1988;69(2):225–244.

[104] White AD. Visual-motor learning in minimally invasive surgery . Leeds, UK; 2016.

[105] Rickard TC. Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. J Exp Psychol. 1997;126(3):288–311.

[106] Bakan D. A generalization of Sidman's results on group and individual functions, and a criterion. Psychol Bull. 1954;51(1):63–64.

[107] Estes WK. The problem of inference from curves based on group data. Psychol Bull. 1956;53(2):134–140.

[108] Sidman M. A note on functional relations obtained from group data. Psychol Bull. 1952;49(3):263–269.

[109] Estes W. Traps in the route to models of memory and decision. Psychol Bull Rev. 2002;9:3–25.

[110] Brown S, Heathcote A. Averaging learning curves across and within participants. Behav Res Methods Instrum Comput. 2003;35(1):11–21.

[111] Nguyen N. Review: Perceiving talking faces: From speech perception to a behavioral principle by Massaro, D. W., MIT Press, Cambridge, Massachusetts: 1998. J Phon. 2000;28:103–109.

[112] Anderson R; 1999. Personal Communication.

[113] Newell KM, Liu YT, Mayer-Kress G. Time scales in motor learning and development. Psychol Rev. 2001;108(1):57–82.

[114] Raw RK. The effect of old age on motor control: Performance and learning. Institute of Psychological Sciences, The University of Leeds; 2012.

[115] Rogers DA, Regehr G, Howdieshell TR, Yeh KA, Palm E. The impact of external feedback on computer-assisted learning for surgical technical skill training. Am J Surg. 2000;179(4):341–343.

[116] Xeroulis GJ, Park J, Moulton CA, Reznick RK, LeBlanc V, Dubrowski A. Teaching suturing and knot-tying skills to medical students: a randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. Surgery. 2007;141(4):442–449.

[117] Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA. 2011;306(9):978–988.

[118] Stefanidis D, Korndorffer JR, Heniford BT, Scott DJ. Limited feedback and video tutorials optimize learning and resource utilization during laparoscopic simulator training. Surgery. 2007;142(2):202–206.

[119] Lim TO, Soraya A, Ding LM, Morad Z. Assessing doctors' competence: application of CUSUM technique in monitoring doctors' performance. Int J Qual Health Care. 2002;14(3):251–258.

[120] Søvik TT, Aasheim ET, Kristinsson J, Schou CF, Diep LM, Nesbakken A, et al. Establishing laparoscopic Roux-en-Y gastric bypass: perioperative outcome and characteristics of the learning curve. Obes Surg. 2009;19(2):158–165.

[121] Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev. 2009;(1):1–39.

[122] Loftus GR, Masson ME. Using confidence intervals in within-subject designs. Psychon Bull Rev. 1994;1(4):476–490.

[123] Bakeman R, Mcarthur D. Picturing repeated measures: Comments on Loftus, Morrison, and others. Behav Res Methods Instrum Comput. 1996;28(4):584–589.

[124] Cousineau D. Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. Tutor Quant Methods Psychol. 2005;1(1):42–45.

[125] Morey RD. Confidence intervals from normalized data: A correction to Cousineau (2005). Tutor Quant Methods Psychol. 2008;4(2):61–64.

[126] Morrison GR, Weaver B. Exactly how many p values is a picture worth? A commentary on Loftus's plot-plus-error-bar approach. Behav Res Methods Instrum Comput. 1995;27(1):52–56.

[127] Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39(2):175–191.

[128] Cave KR, Wolfe JM. Modeling the role of parallel processing in visual search. Cogn Psychol. 1990;22(2):225–271.

[129] Treisman A. Features and objects: The fourteenth Bartlett memorial lecture. Q J Exp Psychol. 1988;40(2):201–237.

[130] Lamy D, Tsal Y. Object features, object locations, and object files: Which does selective attention activate and when? J Exp Psychol. 2000;26(4):1387–1400.

[131] Tsal Y, Lavie N. Location dominance in attending to color and shape. J Exp Psychol Hum Percept Perform. 1993;19(1):131–139.

[132] Liu T, Stevens ST, Carrasco M. Comparing the time course and efficacy of spatial and feature-based attention. Vision Res. 2007;47(1):108–113.

[133] Guérard K, Morey CC, Lagacé S, Tremblay S. Asymmetric binding in serial memory for verbal and spatial information. Mem Cognit. 2013;41(3):378–391.

[134] Chalfonte BL, Johnson MK. Feature memory and binding in young and older adults. Mem Cognit. 1996;24(4):403–416.

[135] Logie R, Engelkamp J, Dehn J, Rudkin S. Actions, mental actions, and working memory. In: Denis M, Logie R, Cornoldi C, De Vega M, Englekamp J, editors. Imagery, Language and Visuo-Spatial Thinking. Current Issues in Thinking and Reasoning. Psychology Press; 2000. p. 161–184.

[136] Campo P, Poch C, Parmentier FBR, Moratti S, Elsley JV, Castellanos NP, et al. Oscillatory activity in prefrontal and posterior regions during implicit letter-location binding. Neuroimage. 2010;49(3):2807–2815.

[137] Jiang Y, Olson IR, Chun MM. Organization of visual short-term memory. J Exp Psychol Learn Mem Cogn. 2000;26(3):683–702.

[138] Brewer CA, Hatchard GW, Harrower MA. ColorBrewer in Print: A Catalog of Color Schemes for Maps. Cartogr Geogr Inf Sci. 2003;30(1):5–32.

[139] Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In: Selected Papers of Hirotugu Akaike. New York, NY: Springer New York; 1998. p. 199–213.

[140] Burnham KP, Anderson DR. Model selection and multimodel inference: A practical information-theoretic approach. 2nd ed. New York: Springer-Verlag; 2002.

[141] Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat. 1951;22(1):79–86.

[142] Symonds MRE, Moussalli A. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behav Ecol Sociobiol. 2010;65(1):13–21.

[143] Levene H. Robust tests for equality of variances. In: Olkin I, editor. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press; 1960. p. 278–292.

[144] Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE. Attentional modulation of neural processing of shape, color, and velocity in humans. Science. 1990;248(4962):1556–1559.

[145] Adatepe M, Adibelli ZH, Esen OS, Imamoglu C, Yildirim M, Erkan N. Anatomic variations of biliary ducts: Magnetic resonance cholangiopancreatography findings of 1041 consecutive patients. Eur Surg. 2016;48(5):296–303.

[146] Lamah M, Karanjia ND, Dickson GH. Anatomical variations of the extrahepatic biliary tree: review of the world literature. Clin Anat. 2001;14:167–172.

[147] Zadeh LA. Fuzzy sets. Inform Control. 1965;8(3):338–353.

[148] Jordan JA, Gallagher AG, McGuigan J, McGlade K, McClure N. A comparison between randomly alternating imaging, normal laparoscopic imaging, and virtual reality training in laparoscopic psychomotor skill acquisition. Am J Surg. 2000;180(3):208–211.

[149] Braun DA, Aertsen A, Wolpert DM, Mehring C. Motor task variation induces structural learning. Curr Biol. 2009;19(4):352–357.

[150] Turnham EJA, Braun DA, Wolpert DM. Facilitation of learning induced by both random and gradual visuomotor task variation. J Neurophysiol. 2012;107(4):1111–1122.

[151] White AD, Mushtaq F, Giles O, Wood ML, Mole C, Culmer PR, et al. Laparoscopic motor learning and workspace exploration. J Surg Educ. 2016;73(6):992–998.

[152] Nisar A. Re: Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ. 2009;338(may14 2):b1802–b1802.

[153] Hsu J. Multiple comparisons: Theory and methods. Chapman & Hall; 1996.

[154] Ruthruff E, Remington RW, Johnston JC. Switching between simple cognitive tasks: the interaction of top-down and bottom-up factors. J Exp Psychol Hum Percept Perform. 2001;27(6):1404–1419.

[155] Jersild AT. Mental set and shift. Arch Psychol. 1927;89:1–81.

[156] Spector A, Biederman I. Mental set and mental shift revisited. Am J Psychol. 1976;89(4):669.

[157] Allport DA, Styles EA, Hsieh S. Shifting intentional set: Exploring the dynamic control of tasks. In: Umiltà C, Moscovitch M, editors. Attention and performance 15: Conscious and nonconscious information processing. MIT Press; 1994. p. 421–452.

[158] Waszak F, Hommel B, Allport A. Task-switching and long-term priming: Role of episodic stimulus–task bindings in task-shift costs. Cogn Psychol. 2003;46:361–413.

[159] Rubin O, Meiran N. On the origins of the task mixing cost in the cuing task-switching paradigm. J Exp Psychol Learn Mem Cogn. 2005;31(6):1477–1491.

[160] Garner WR. The stimulus in information processing. Am Psychol. 1970;25(4):350–358.

[161] Garner W. The Processing of Information and Structure. Psychology Press; 1974.

[162] de Jong R. Strategical determinants of compatibility effects with task uncertainty. Acta Psychol. 1995;88:187–207.

[163] Meuter R, Allport A. Bilingual language switching in naming: Asymmetrical costs of language selection. J Mem Lang. 1999;40(1):25–40.

[164] Allport D, Wylie G. Task-switching: Positive and negative priming of task-set. In: Humphreys G, Duncan J, AM T, editors. Attention, space and action: Studies in cognitive neuroscience. Oxford University Press; 1999. p. 273–296.

[165] Allport D, Wylie G. "Task-switching" stimulus–response bindings, and negative priming. In: Monsell J, Driver J, editors. Control of cognitive processes: Attention and performance XVIII. MIT Press; 2000. p. 35–70.

[166] Yeung N, Monsell S. Switching between tasks of unequal familiarity: the role of stimulus-attribute and response-set selection. J Exp Psychol. 2003;29(2):455–469.

[167] Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol. 1935;18(6):643–662.

[168] MacLeod CM, Dunbar K. Training and Stroop-like interference: evidence for a continuum of automaticity. J Exp Psychol Learn Mem Cogn. 1988;14(1):126–135.

[169] Cohen JD, Dunbar K, McClelland JL. On the control of automatic processes: a parallel distributed processing account of the Stroop effect. Psychol Rev. 1990;97(3):332–361.

[170] Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. Int J Inform Manage. 2015;35(2):137–144.

[171] Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA. 2013;309(13):1351–1352.

[172] Nouraei SAR, Hudovsky A, Frampton AE, Mufti U, White NB, Wathen CG, et al. A study of clinical coding accuracy in surgery. Ann Surg. 2015;261(6):1096–1107.

[173] Bohnen JD, Chang DC, Lillemoe KD. Reconceiving the morbidity and mortality conference in an era of big data. Ann Surg. 2016;263(5):857–859.

[174] Mathias B, Lipori G, Moldawer LL, Efron PA. Integrating "big data" into surgical practice. Surgery. 2016;159(2):371–374.

[175] Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015;372(9):793–795.

[176] Carter P. Operational productivity and performance in English NHS acute hospitals: Unwarranted variations; 2016. Department of Health.

[177] Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. JAMA. 2014;311(24):1–2.

[178] Mann RP, Mushtaq F, White AD, Mata-Cervantes G, Pike TW, Coker D, et al. The problem with big data: Operating on smaller datasets to bridge the implementation gap. Front Public Health. 2016;4:1–4.

[179] Dunning T. Natural experiments in the social sciences: A design-based approach. Cambridge University Press; 2012.

[180] Propper C. The demand for private health care in the UK. J Health Econ. 2000;19(6):855–876.

[181] Mulvaney C, Coupland C, Wilson A, Hammersley V, Dyas J, Carlisle R. Does increased use of private health care reduce the demand for NHS care? A prospective survey of general practice referrals. J Public Health (Oxf). 2005;27(2):182–188.

[182] National Audit Office. Review of the final benefits statement for programmes previously managed under the National Programme for IT in the NHS. 2013;p. 1–24.

[183] House of Commons Committee of Public Accounts. The dismantled National Programme for IT in the NHS. 2013;p. 1–43.

[184] Syal R. Abandoned NHS IT system has cost £10bn so far. The Guardian. 2013;p. 1–3.

[185] Figueras J, McKee M, Mossialos E, Saltman RB. Social health insurance systems in western Europe. New York, NY, USA: European Observatory on Health Systems and Policies; 2004.

[186] American Society of Anesthesiologists. New Classification of Physical Status. Anesthesiology. 1963;24:111.

[187] Monsell S. Task switching. Trends Cogn Sci. 2003;7(3):134–140.

[188] AXA PPP Healthcare. Specialist Procedure Codes; 2017.

[189] Garnett GM, Wilkens L, Limm W, Wong LL. Operative time as a measure of quality in pancreaticoduodenectomy: Is faster better? A retrospective review using the ACS NSQIP database. Surg Sci. 2015;6:418–426.

[190] Harrison OJ, Smart NJ, White P, Brigic A, Carlisle ER, Allison AS, et al. Operative time and outcome of enhanced recovery after surgery after laparoscopic colorectal surgery. JSLS. 2014;18(2):265–272.

[191] Jackson TD, Wannares JJ, Lancaster RT, Rattner DW, Hutter MM. Does speed matter? The impact of operative time on outcome in laparoscopic surgery. Surg Endosc. 2011;25(7):2288–2295.

[192] Colman M, Wright A, Gruen G, Siska P, Pape HC, Tarkin I. Prolonged operative time increases infection rate in tibial plateau fractures. Injury. 2013;44(2):249–252.

[193] Strum DP, Sampson AR, May JH, Vargas LG. Surgeon and type of anesthesia predict variability in surgical procedure times. Anesthesiology. 2000;92(5):1454–1466.

[194] Vacanti CJ, VanHouten RJC, Hill RC. A statistical analysis of the relationship of physical status to postoperative mortality in 68,388 cases. Anesth Analg. 1970;49(4):564–566.

[195] Farrow SC, Fowkes FG, Lunn JN, Robertson IB, Samuel P. Epidemiology in anaesthesia. II: Factors affecting mortality in hospital. Br J Anaesth. 1982;54(8):811–817.

[196] Pedersen T, Eliasen K, Ravnborg M, Viby-Mogensen J, Qvist J, Johansen SH, et al. Risk factors, complications and outcome in anaesthesia. A pilot study. Eur J Anaesthesiol. 1986;3(3):225–239.

[197] Wolters U, Wolf T, Stützer H, Schröder T. ASA classification and perioperative variables as predictors of postoperative outcome. Br J Anaesth. 1996;77(2):217–222.

[198] Alexander KP, Anstrom KJ, Muhlbaier LH, Grosswald RD, Smith PK, Jones RH, et al. Outcomes of cardiac surgery in patients age $\geq$80 years: results from the National Cardiovascular Network. J Am Coll Cardiol. 2016;35(3):1–8.

[199] Latkauskas T, Rudinskaitė G, Kurtinaitis J, Jančiauskienė R, Tamelis A, Saladžinskas Ž, et al. The impact of age on post-operative outcomes of colorectal cancer patients undergoing surgical treatment. BMC Cancer. 2005;5(1):177–5.

[200] Al-Refaie WB, Parsons HM, Habermann EB, Kwaan M, Spencer MP, Henderson WG, et al. Operative Outcomes Beyond 30-day Mortality. Ann Surg. 2011;253(5):947–952.

[201] Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Soft. 2015;67(1):1–48.

[202] Schutte S, Donnay K. Matched wake analysis: finding causal relationships in spatiotemporal event data. Polit Geogr. 2014;41:1–10.

[203] Monsell S, Sumner P, Waters H. Task-set reconfiguration with predictable and unpredictable task switches. Mem Cognit. 2003;31(3):327–342.

[204] Nieuwenhuis S, Monsell S. Residual costs in task switching: testing the failure-to-engage hypothesis. Psychon Bull Rev. 2002;9(1):86–92.

[205] De Jong R. An intention-activation account of residual switch 43 costs. In: Monsell S, Driver J, editors. Control of Cognitive Processes: Attention and Performance. MIT Press; 2000. p. 357–376.

[206] White AD, Skelton M, Mushtaq F, Pike TW, Mon-Williams M, Lodge J, et al. Inconsistent reporting of minimally invasive surgery errors. Ann R Coll Surg Engl. 2015;97(8):608–612.

[207] Regenbogen SE, Greenberg CC, Studdert DM, Lipsitz SR, Zinner MJ, Gawande AA. Patterns of technical error among surgical malpractice claims: an analysis of strategies to prevent injury to surgical patients. Ann Surg. 2007;246(5):705–711.

[208] Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. BMJ. 2001;322(7285):517–519.

[209] Sohn MH, Ursu S, Anderson JR, Stenger VA, Carter CS. The role of prefrontal cortex and posterior parietal cortex in task switching. Proc Natl Acad Sci USA. 2000;97(24):13448–13453.

[210] Kimberg DY, Aguirre GK, D'Esposito M. Modulation of task-related neural activity in task-switching: an fMRI study. Brain Res Cogn Brain Res. 2000;10(1-2):189–196.

[211] Rogers RD, Monsell S. Costs of a predictible switch between simple cognitive tasks. J Exp Psychol. 1995;124(2):207–231.

[212] Stafford T, Dewar M. Tracing the trajectory of skill learning with a very large sample of online game players. Psychol Sci. 2014;25(2):511–518.

[213] Fletcher D, Edwards D, Tolchard S, Baker R, Berstock J. Improving theatre turnaround time. BMJ Qual Improv Report. 2017;6(1):u219831.w8131–6.

[214] Reitsma AM, Moreno JD. Ethical regulations for innovative surgery: the last frontier? JACS. 2002;194(6):792–801.

[215] Miskovic D. Proficiency gain and competency assessment in laparoscopic colorectal surgery. Department of Surgery and Cancer. Imperial College London; 2012.

[216] King B, Jatoi I. The mobile Army surgical hospital (MASH): a military and surgical legacy. J Natl Med Assoc. 2005;97(5):648–656.

[217] Pike TW, Stobbs N, Mushtaq F, Lodge JPA. The effects of an e-textbook and the 'reverse classroom' on surgical training. Ann R Coll Surg Engl (Suppl). 2015;p. 10–13.

[218] Pike TW, Mushtaq F, Wilkie RM, Lodge J, Mon-Williams M. How should surgeons warm up? An experimental psychology approach. JSS. 2017;4:B6–B6.

[219] Pike TW, Mushtaq F, Mann RP, Chambers P, Hall G, Tomlinson JE, et al. Operating list composition and surgical performance. Br J Surg. 2018;.