ORIGINAL STUDY

''Contemplating the Next Maneuver''

Functional Neuroimaging Reveals Intraoperative Decision-making Strategies

Daniel Richard Leff, FRCS, PhD,† Gabriella Yongue, MBBS, BSc,*† Ivo Vlaev, PhD,†*
Felipe Orihuela-Espina, PhD,§ David James, MRCS, PhD,*† Michael J. Taylor, PhD,†*
Thanos Athanasiou, PhD, FECTS,† Ray Dolan, MD, FMedSc, FRS,‡
Guang-Zhong Yang, FREng, PhD, and Ara Darzi, MD, FRCS, FACS, FMedSci, FRS*†*

Objective: To investigate differences in the quality, confidence, and consistency of intraoperative surgical decision making (DM) and decision systems AQ4 that operators use using functional neuroimaging.

Summary Background Data: Novices are hypothesized to use conscious analysis (effortful DM) leading to activation across the dorsolateral prefrontal cortex, whereas experts are expected to use unconscious automation (habitual DM) in which decisions are recognition-primed and prefrontal cortex independent.

Methods: A total of 22 subjects (10 medical student novices, 7 residents, and 5 attendings) reviewed simulated laparoscopic cholecystectomy videos, determined the next safest operative maneuver upon video termination (10 s), and reported decision confidence. Video paradigms either declared (''primed'') or withheld (''unprimed'') the next operative maneuver. Simultaneously, changes in cortical oxygenated hemoglobin and deoxygenated hemoglobin inferring prefrontal activation were recorded using Optical Topography. Decision confidence, consistency (primed vs unprimed), and quality (script concordance) were assessed.

Results: Attendings and residents were significantly more certain ($P < 0.001$), and decision quality was superior (script concordance: attendings = 90%, residents = 78.3%, and novices = 53.3%). Decision consistency was significantly superior in experts ($P < 0.001$) and residents ($P < 0.05$) than novices ($P = 0.183$). During unprimed DM, novices showed significant activation of the dorsolateral prefrontal cortex, whereas this activation pattern was not observed among residents and attendings. During primed DM, significant activation was not observed in any group.

Conclusions: Expert DM is characterized by improved quality, consistency, and confidence. The findings imply attendings use a habitual decision system, whereas novices use an effortful approach under uncertainty. In the presence of operative cues (primes), novices *disengage* the prefrontal cortex and seem to accept the observed operative decision as correct.

Keywords: brain, decision making, functional near-infrared spectroscopy, prefrontal, simulation, surgery, training

(*Ann Surg* 2016;xx:xxx–xxx)

A surgeon's ability to make reasoned judgments under pressure during operative interventions influences surgical workflow and patient safety. Accurate perception and interpretation of the dynamic nature of the operative scene known as situational awareness (SA)[1] and appropriate decision making (DM) to guide sequential operative maneuvers should be considered safety-critical skills. Yet, although there has been a systematic focus on training and assessment of technical skills, research pertaining to surgical cognition in general[2] and operative SA[3] or DM in particular[4] are scant, possibly due to the challenges associated with investigating complex executive functions.[5]

Operative DM can be simplified as a continuous cycle of monitoring and SA, appropriate action taking, and outcome evaluation to update and improve the operator's DM system.[5] As shown in Figure 1, within this model exist a range of DM strategies that can be actioned depending upon the available time, perceived risk to the patient, and experience of the operator. For example, expert surgeons encountering a familiar operative scene are anticipated to engage a *recognition-primed* approach to select solutions from memory. Conversely, residents with limited domain experience are hypothesized to associate operative scenes with a set of action rules known as ''habit learning'' (or habitual DM which involves learning the value of actions in different states of the world), or to use analytical DM to compare and contrast the perceived risks, associated with a range of possible solutions (eg, ''dissect'' vs ''divide''), known as ''goal-directed learning'' (or goal-directed DM which involves explicit knowledge of the action–outcome contingencies).[6,7] Furthermore, for the expert trainer guiding a resident through an intervention, SA also involves assessments of the trainee's DM system, allowing the procedure to flow where trainer–resident DM seems congruent but importantly knowing when to veto incorrect decisions and take back control. The latter often relies on an incongruent behavioral trigger or cue such as the resident inserting a pair of scissors when the trainer perceives that more dissection is required. Experimentally, surgical simulation facilitates manipulation of behavioral cues, which can be covertly introduced as an ''unconscious prime'' to investigate the impact they may have on trainer DM.

Critically, expertise in operative DM is unlikely to be revealed in behavioral responses such as action selection or choice of operative maneuvers per se because the internal rumination of ''what to do next'' in surgery does not have a behavioral correlate that can be linearly mapped. Instead, we anticipate that disparities in intraoperative DM

AQ1 AQ2 From the *Hamlyn Centre for Robotic Surgery, Imperial College London, London, UK; †Department of BioSurgery and Surgical Technology, Imperial College London, London, UK; ‡Welcome Trust Centre for NeuroImaging, University College London, London, UK; and §Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, Mexico.

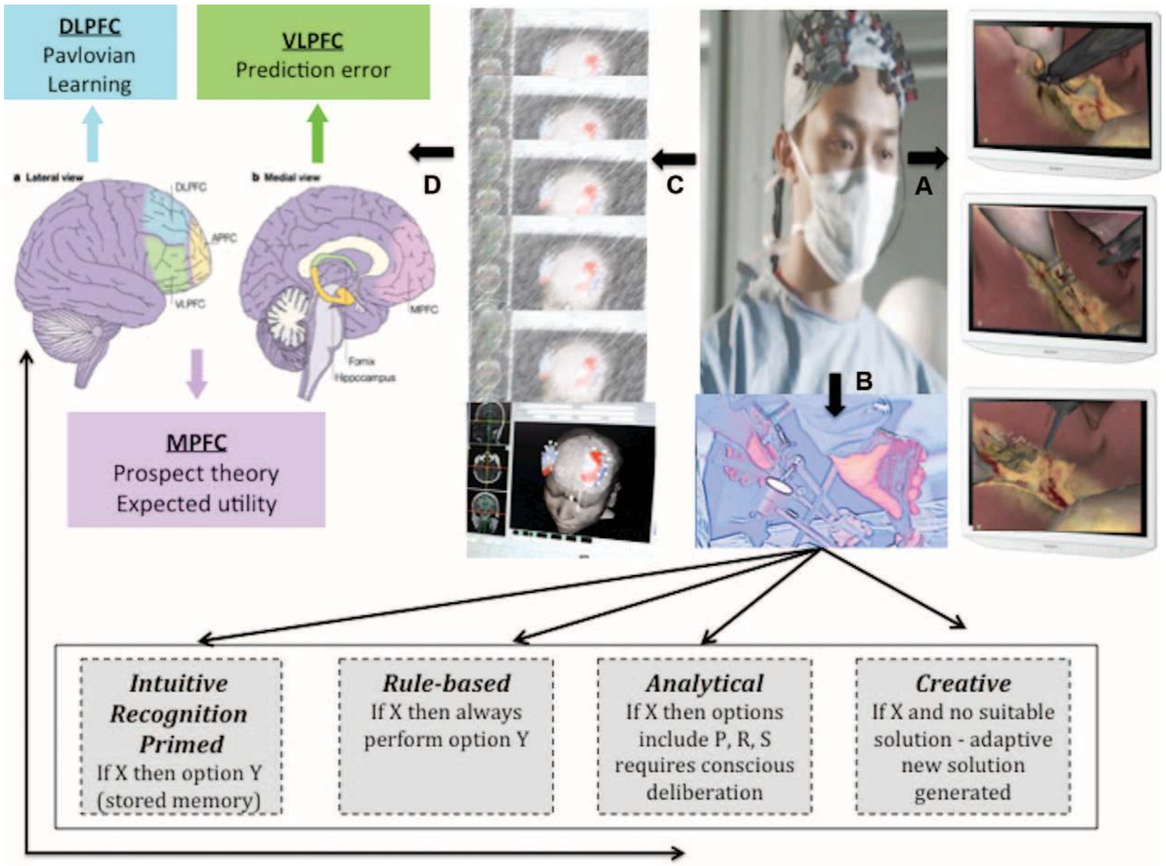
Authors' contributions: The study design was conceived and developed by DRL, GY, IV, RD, G-ZY, and AD. DRL, DJ, and IV executed the experiment and collected the functional imaging data. Data preprocessing and statistical analysis were conducted by DRL, GY, FO-E, MT, and TA. Data interpretation was performed by DRL, FO-E, and IV in consultation with RD, G-ZY, and AD. The manuscript was drafted by DRL, GY, and IV. Critical editing of the manuscript was performed by RD, G-ZY, and AD.

AQ5 Disclosure: The authors report no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

AQ3 Reprints: Ara Darzi, MD, FRCS, FACS, FMedSci, FRS, Hamlyn Chair of Surgery, Academic Surgical Unit, 10th floor, QEQM Wing, St Mary's Hospital, London W2 1NY, UK. E-mail: a.darzi@imperial.ac.uk

ISSN: 0003-4932/14/26105-0821
DOI: 10.1097/SLA.0000000000001651



AQ10 **FIGURE 1.** A proposed 2-step model of surgeons' intraoperative decision making, adapted from Flin et al[5] to incorporate a research hypothesis based on intraoperative neuromonitoring. Surgeons closely monitor the operative scene (A), assess the operative anatomy, and use an appropriate DM strategy (B) to select the next safest operative maneuver. The strategy used depends on available time, perceived risk, and operator experience. The hypothesis is that experts use a recognition-primed approach, whereas novices ruminate options using an analytical DM strategy. Within a neuroimaging framework, surgeons are monitored with multichannel OT such that at each DM phase optical brain data are acquired, and subsequently processed and analyzed to determine the loci of greatest response from which the DM system used can be elucidated (C). Analytical DM evokes dorsolateral prefrontal (DLPFC—operant learning), ventrolateral prefrontal (VLPFC—prediction errors), and medial prefrontal activations (MPFC—prospect theory and expected utility) (D).

manifest as differences in the internal decision systems and cognitive strategies operators use. Therefore, the scientific challenge is how to reliably interrogate surgeons to unveil operative DM strategy. This is important given that intraoperative errors are more commonly due to errors in perception, judgment, and DM,[8,9] and that errors in surgery persist despite significant efforts to improve skills training during residency. Bile duct injuries during laparoscopic cholecystectomy, for example, have cost an estimated $33 million in medico-legal claims in the United Kingdom[10]and $214,000 per claimant in the United States.[11] Bile duct injuries are more commonly due to unconscious assumptions and optical illusions,[8] or failure to establish a "critical view of safety" leading to errors in DM.[9] Moreover, despite recent calls for assessment of attention and concentration,[12] and operator perception of decision risk,[13] there has been no systematic approach to assess surgeons' cognitions intraoperatively. Although postevent interviewing of surgeons provides a degree of insight, the approach is time-consuming, subjective, and cannot be used to anchor residents' progress through training.[4,14,15] An alternative strategy is to capitalize on developments in noninvasive functional neuroimaging technologies to monitor operator brain function during operative interventions on the basis that the magnitude or pattern of cortical response correlates with the decision system used.

The brain contains multiple distinct decision systems,[6,7] differentiated according to their engagement of the corticostriatolimbic circuits in the brain.[16] Each system assigns a "value" to available actions, and thus competes with the actions favored by other systems.[17] Recent evidence indicates competition between a cognitive, goal-directed planning system centered in the lateral prefrontal cortex and parietal cortex, and habitual decision system associated with dopamine and the basal ganglia.[18,19] Decisions requiring effort, working memory, and deductive reasoning have been shown to activate the dorsolateral prefrontal cortex (DLPFC),[20,21] whereas habitual decisions are stimulus-response associations learned through repeated practice and rewards in a stable environment (such mental habits are usually the consequence of past goal pursuits, but once acquired, habits are cued and performed without mediation of a goal).[22] As one's experience accumulates, control over decisions gradually transfers from goal-directed process, which demands effort and time, to the habitual processes which are rapid and easy to execute.[7] On the basis of this evidence and DM theories already outlined, novice surgeons are expected to recruit the DLPFC to a greater extent than expert surgeons owing to escalated levels of uncertainty, need for internal cross-referencing, and more detailed analysis of options during operative DM.

METHODS

Subjects

After local regional ethical approval (LREC: 05/Q0403/142), 22 healthy individuals were recruited from Imperial College London and Imperial College Healthcare NHS Trust. Participants were subdivided into 3 groups according to prior operative expertise in laparoscopic cholecystectomy as follows: 10 medical students [mean age ± SD (yrs) = 22.40 ± 0.97] with no prior experience of laparoscopy were classified as ''novices''; 7 participants were ''residents'' enrolled in specialty training schemes [mean age ± SD (yrs) 32.14 ± 1.77] and had prior experience of assisting on laparoscopic cholecystectomy or performing the procedure under supervision (see Supplementary Table 1, http://links.lww.com/SLA/A972); and 5 attendings were classified as ''experts'' [mean age ± SD (yrs) = 32.14 ± 1.77] on the basis of more than 100 independent laparoscopic cholecystectomies. A history of neuropsychiatric disorders was an exclusion criterion (n = 0), and all participants were asked to refrain from alcohol and caffeine for 24 hours given the known effects on cerebral hemodynamics.[23]

Task and Training

Before the experiment, all subjects were provided with a training session that included an overview of the operative anatomy, principles, and operative steps of laparoscopic cholecystectomy (ie, Calot's triangle dissection, critical view of safety, clipping of cystic artery and duct, etc). After training, subjects were asked to complete a short test that posed questions to evaluate knowledge and understanding of the operative anatomy and procedural flow of laparoscopic cholecystectomy (see questionnaire supplementary content, http://links.lww.com/SLA/A972). Failure to achieve perfect score in the test led to exclusion (n = 0). After successful test completion, subjects proceeded to the DM experiment.

Operative DM Paradigm Experimental Set-up

The experiment focused on interrogating intraoperative DM during laparoscopic cholecystectomy. Subjects were asked to regard a monitor and observe a series of video clips (n = 12) of high-fidelity-simulated laparoscopic cholecystectomy (prerecorded using LapMentor, Simbionix, Israel). Each video clip lasted 10 seconds, revealed a sequence of operative maneuvers at random (ie, unpredictable), and terminated at a point at which an operative decision was required. Video clips were classified as either *''primed''* (n = 5) in which the operator's next step was readily declared (eg, scissors brought into view suggesting DM to cut), or *''unprimed''* (n = 7) which terminated immediately after a given action without indication of what occurred next in the simulation (Fig. 2A). The sequence in which subjects experienced primed and unprimed video clips was randomized. After each video clip, subjects were asked to verbally report the recommended next operative maneuver from a list provided on the monitor. Each operative decision was recorded by the investigators (DRL, DJ, and GY). After the DM task, subjects were asked to state how confident they were of their decision on a scale of 1 to 6 (1 = low confidence, 6 = high confidence).

Experimental Set-up and Block Design Experiment

As shown in Figure 2B, a block design experiment was conducted comprising 12 sequential blocks, each comprising episodes of ''rest,'' and 3 *stimuli* identified as ''video review,'' ''decision,'' and ''confidence.'' During rest periods (30 s) subjects were seated and asked to place their hands on a table and focus on a fixation cross. During video review subjects were instructed to pay close attention to the operative video clip (10 s) with a view to reporting the next operative maneuver upon video termination. During decision episodes a slide was presented as an aide-memoire of the surgical options (eg, dissect further, divide cystic artery, convert to open, etc) and subjects verbally reported their decision (10 s). Finally, subjects reported decision confidence (10 s). Before progression to the next video clip, a posttrial rest period (30 s) was introduced to enable cortical hemodynamics to return to baseline. Cortical activity was measured throughout using functional near-infrared spectroscopy-based Optical Topography (OT), which converts changes in light levels into changes in cortical hemodynamics[24] and therefore monitors the hemodynamic response to neuronal activation (''neurovascular coupling principle'').[25] The typical hemodynamic response to neuronal activation comprises a rise in oxygenated hemoglobin (HbO_2) and a decrease in deoxygenated hemoglobin (HHb).

Functional Neuroimaging

Subjects were neuromonitored using a commercial OT system (ETG-4000; Hitachi Medical Corp, Tokyo, Japan). OT is a portable, noninvasive technique that is resistant to motion artifact and has been successfully used in the study of technical skills in the field of surgery.[2] Multichannel OT is a technique that measures changes in light levels across multiple cortical locations simultaneously. Light is shone on to the subject's scalp (700–900 nm), and attenuated light is detected by neighboring photodiode detectors. The modified Beer-Lambert Law[26] was used to compute relative changes in hemoglobin concentration at multiple locations between emitters and detectors (referred to as ''channels''). Here, 15 optodes (emitters/detectors) were deployed 30 mm apart in a 5 × 3 flexible plastic array positioned according to the 10 to 20 system of electrode placement to monitor hemodynamic change across the PFC,[27] as shown in Figure 2B. Near-infrared light at 695 and 830 nm was emitted from 8 optical fiber sources and detected by 7 neighboring avalanche photodiode detectors, resulting in 22 different measuring channels. Probes were fastened into C-shaped metallic holders, and the entire array was secured to the operator's scalp using surgical bandage (Surgifix, Colorline, Italy), as shown in Figure 2B.

Stress

Subjective levels of stress were monitored on the basis that stress-related changes in systemic physiology might influence functional OT data.[28] Subjects were asked to complete short form of the Spielberger State-Trait Anxiety Inventory (STAI) before, during, and after the study.

Data Processing and Statistical Analysis

Decision Quality, Consistency, and Confidence

The quality of DM responses was assessed using script concordance, which is a tool designed to assess clinical reasoning on the basis that judgment can be probed and concordance with a reference panel of experts measured.[29] Script concordance is calculated by scoring each decision by comparing it with the DM of a panel of expert surgeons. Here, we invited a panel of expert consultant surgeons not recruited to the study (n = 10) to review each

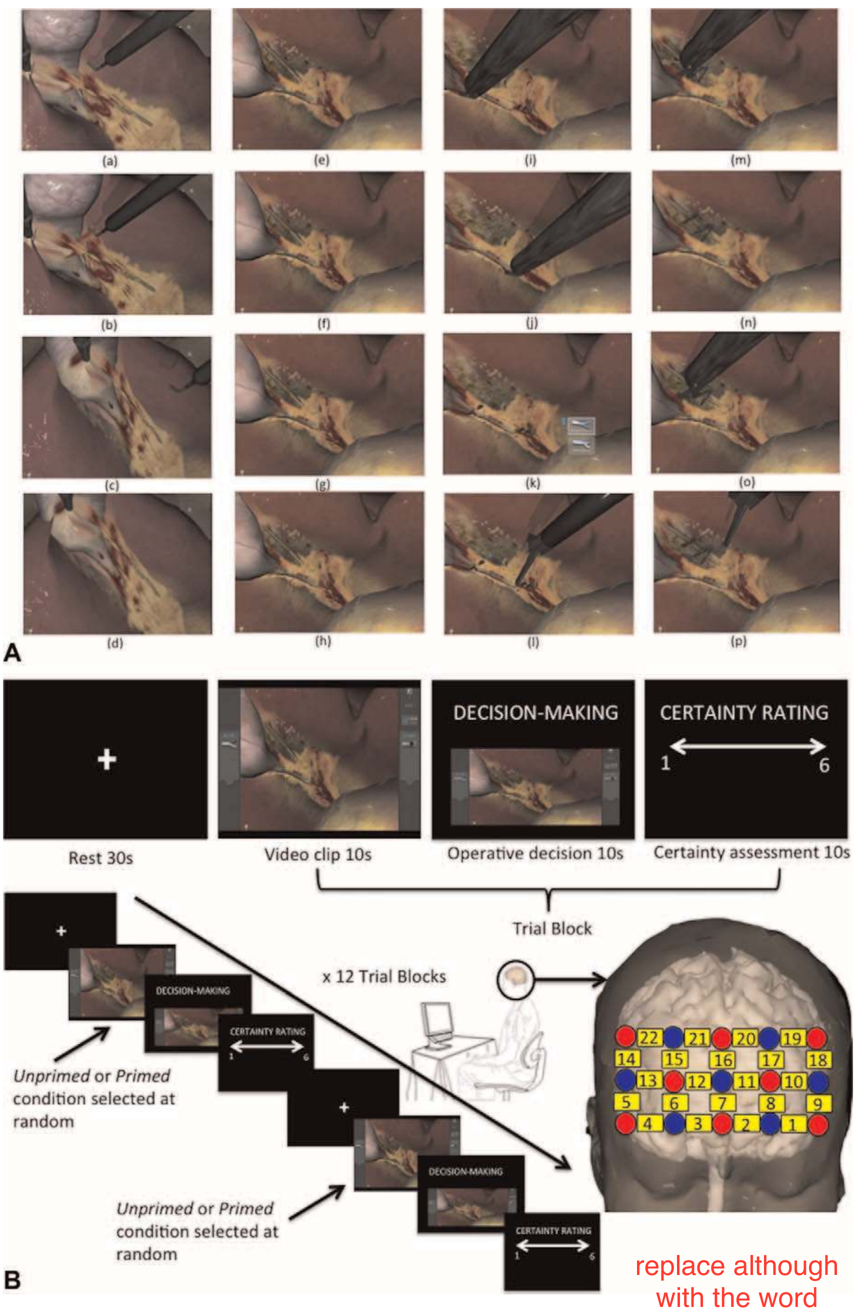


FIGURE 2. (Panels A–B): A, Images depicting different phases of simulated laparoscopic cholecystectomy. Videos were classified as either *unprimed* (eg, a–d and e–h) that terminated at a point where the operator's next maneuver was not apparent (d/h) or *primed* (eg, i–l and n–p) that revealed the operator's intention, for example, to clip or divide a structure (l/p). Examples of *unprimed* videos include episodes of Calot's triangle dissection (a–d) or gallbladder manipulation without dissection (e–h), after which further dissection would be required in both cases before cystic duct and artery could be safely clipped and ligated. Examples of *primed* videos include sequences of clipping and dividing the cystic duct (I–l) or the cystic artery (n–p). At termination of these primed video sequences, the operator's decision to divide the structure is both clear and incorrect (ie, clips placed too low down near the common bile duct (I–l), and clipping of the cystic duct should proceed division of the cystic artery (n–p). B, Experimental task set up. Subjects were seated at a table and observed video sequences of simulated laparoscopic cholecystectomy. The experiment was delivered as a block design, with repeated episodes of rest (30 s) interspersed with trial blocks that were composed of 3 substimuli, namely video clip review (10 s), operative decision making (10 s), and confidence ratings (10 s). During rest periods subjects observed the fixation cross, during video review they observed a certain phase of laparoscopic cholecystectomy, and during decision making trials they viewed the video's final image and were asked to report the next safest operative maneuver. Finally, they were asked to report their confidence in decision making. Video clips were classified either *primed* or *unprimed* as to whether the operator's next move was declarative or not. The sequence to which subjects were exposed to these 2 conditions was random. In total, subjects were exposed to 12 trial blocks, ~~although~~ multichannel OT monitored changes in cortical hemodynamic change across 22 channels (yellow numbered squares) positioned across the dorsolateral, ventrolateral, and medial prefrontal cortex.

replace although
with the word
“whilst”

laparoscopic cholecystectomy video used in the experiment and record what was in their expert opinion the correct next operative move. In this regard, we obtained consensus as to the most appropriate next operative step and hence were able to award points for participant DM based on the expert responses (Supplementary Table 2, http://links.lww.com/SLA/A972). Decision consistency was determined by correlating decisions for each “primed” video with the “unprimed” equivalent (10 videos) using Spearman correlation analysis. Decision confidence scores were tabulated according to operator expertise and decision type (ie, “unprimed” and “primed”). The χ^2 test was used to compare confidence between experience groups and also within each experience group comparing “unprimed” and “primed” conditions.

For statistical analysis of decision quality, consistency and confidence $P < 0.05$ was deemed statistically significant.

Functional Neuroimaging Data

Functional neuroimaging data were analyzed using the Imperial College Neuroimaging Analysis, a bespoke software package programmed using Matlab (MathWorks, Natick, MA). Raw optical data were subject to integrity checks to eliminate instrumentation noise, system drift, optode mirroring, and apparent nonrecording, and to increase signal-to-noise ratio.[23] Data were decimated and linearly de-trended, and relative changes in light intensities were converted into changes in hemoglobin concentration using the modified Beer-Lambert Law.[26]

For a given experience group, hemodynamic time courses were produced for each of the 22 channels and visually inspected to identify areas consistent with activation, that is increases in HbO_2 or decreases in HHb, and confirmed using a statistical channel-based analytical framework referred to as the ‘‘activation matrix.’’ Matrices were constructed by assessing task-induced changes in both HbO_2 and HHb. For each channel, average baseline rest Hb data (5 s of data before stimulus onset) were compared with average trial Hb data (17 s of data, 2 s after stimulus onset) using the Wilcoxon sign-rank test. Channels displaying statistically significant ($P < 0.05$) increases in HbO_2 coupled with statistically significant ($P < 0.05$) decreases in HHb were considered *activated*. Conversely, channels displaying the opposing trend were considered *deactivated*. Channels in which directional changes in Hb species were commensurate with either activation or deactivation but for which only 1 Hb species reached statistical threshold were termed ‘‘activation or deactivation trends.’’

Regarding channels displaying activation or activation trends, a new variable termed ‘‘ΔHb’’ was computed to compare the magnitude of cortical hemodynamic change between experience groups. For each channel and Hb species, ΔHb represented the difference between rest Hb data and stimulus Hb data (ie, $\Delta \text{Hb} = \Delta$ D stimulus Hb $- \Delta$ rest Hb). Here, rest data were calculated by averaging the last 5 seconds of each rest period before the video presentation, whereas stimulus data represented the average of 17-second epochs commencing 2 seconds after the stimulus onset. For a given channel, ΔHb data were compared between novices and operators with either prior laparoscopic training or real operative experience (ie, residents and attendings combined) using the Mann-Whitney U test. ΔHb data were further grand averaged across DLPFC channels to obtain individual proxy indicators of brain activity (thus allowing one observation per trial per individual). Finally, a generalized linear mixed model (GLMM) was computed across and within each expertise group, using grand-averaged ΔHb data, with ΔHbO_2 and ΔHHb, as the dependent variable; priming condition (primed vs unprimed) as fixed effects (within-subject factor); and subjects, trial number, and stimulus as random effects.

Stress Data

Within-group comparisons in STAI responses before, during, and after the experiment were analyzed using the Wilcoxon signed-rank test.

RESULTS

Cohort Demographics

Seven female and 15 male subjects participated. No significant sex distribution differences ($\chi^2 = 1.45$, $P = 0.483$) or differences in handedness ($\chi^2 = 5.87$, $P = 0.209$) were identified between the groups. Participant’s ages ranged from 21 to 51 years and experts were significantly older than residents [mean age ± SD (yrs): attendings $= 36.20 \pm 8.79$ vs residents $= 32.14 \pm 1.77$, $P < 0.05$] and novices [mean age ± SD (yrs) = attendings $= 36.20 \pm 8.79$ vs novices $= 22.40 \pm 0.97$, $P < 0.05$].

Operative Decision Confidence

As depicted in Supplementary Figure 1 and Table 3, http://links.lww.com/SLA/A972, DM confidence varied significantly with expertise ($P < 0.001$). A greater proportion of attendings were observed to be highly certain of operative decisions vs residents and novices (% reporting high confidence: attendings = 73%, residents = 60%, novices = 11%). Both attendings and residents were significantly more certain of decisions than novices (mean confidence ± SD: novices $= 3.95 \pm 1.20$, residents $= 5.37 \pm 0.94$, experts $= 5.68 \pm 0.60$; attendings vs novices: $\chi^2 = 87.35$, $P < 0.001$, residents vs novices: $\chi^2 = 71.22$, $P < 0.001$). However, there was no statistical difference in DM confidence between residents and attendings ($\chi^2 = 7.31$, $P = 0.120$). Priming had no significant impact on decision confidence regardless of operator experience (novices: $\chi^2 = 3.60$, $P = 0.730$, residents: $\chi^2 = 2.18$, $P = 0.702$, attendings: $\chi^2 = 1.84$, $P = 0.606$).

Operative Decision Quality, Decision Consistency, and Stress

Script concordance confirmed that attending and resident DM aligned more closely with expert panel decisions [script concordance % (score): attendings = 90 (10.8), residents = 78.3 (9.4), novices = 53.3 (6.4), maximum score = 12]. Attendings more frequently challenged the apparent next operative move in the primed video sequences than did residents or novices [contradict prime decision: attendings = 85.0%, residents = 74.0%, novices = 44.0%]. The frequency with which primed cues were challenged varied significantly with expertise ($\chi^2 = 9.810$, $P = 0.007$). There was a lack of consistency in DM between matched unprimed and primed decision stimuli among novices ($R^2 = 0.191$, $P = 0.183$), whereas residents’ ($R^2 = 0.445$, $P = 0.007$) and attendings’ ($R^2 = 0.524$, $P = 0.001$) responses were significantly more consistent across conditions. There was no statistically significant difference in STAI scores between groups ($P = 0.574$). No significant changes in stress or anxiety were observed across the experiment among residents or attendings (Supplementary Table 4, http://links.lww.com/SLA/A972). However, comparing STAI scores during and after the experiment confirmed a significant decrease in anxiety among novices ($P = 0.011$).

Cortical Hemodynamics

Unprimed Decisions

Activation matrices for unprimed stimuli are illustrated by operator expertise in Figure 3 (panel a) (see supplementary material for full statistical analysis, http://links.lww.com/SLA/A972). Regarding operative video review, a greater number of PFC channels displayed activation trends among novices than residents and attendings (activation trends: novices = 14/12, residents = 4/22, and attendings = 4/22). In addition, although activation was observed across bilateral DLPFC among residents and attendings, activation among novices was predominantly ventromedial in distribution. During DM trials, activated DLPFC channels (ie, statistically significant changes in both HbO_2 and HHb species) were only observed among novices, whereas activation trends were observed across bilateral DLPFC channels among residents and attendings (residents: right DLPFC = 4 channels, left DLPFC = 4 channels; attendings: right DLPFC = 2 channels, left DLPFC = 3 channels). Ventromedial activation trends were observed solely among novices during DM trials.

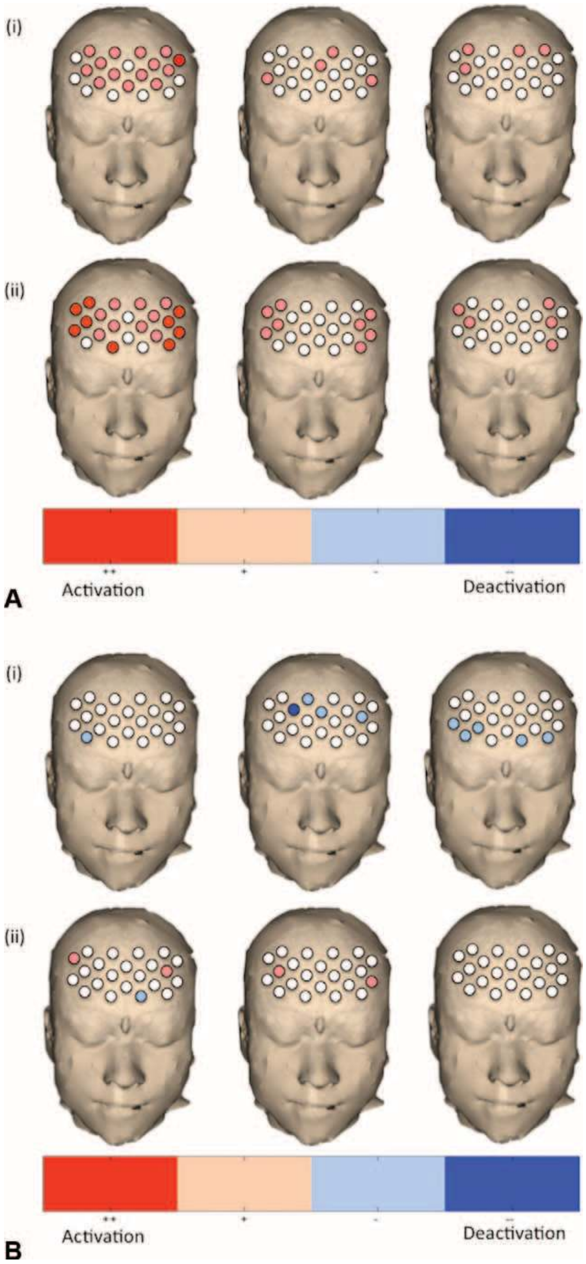


FIGURE 3. (Panels A–B): A, Charts summarize group-averaged statistical analysis of HbO_2 and HHb and presented in the form of series of activation/deactivation matrices. Each plot represents an experience group (left column = novices, middle column = re-residents, right column = attendings) and the *unprimed* conditions either video review (i) or decision-making episodes (ii). Twenty-two channels are highlighted (black circles) and color coded to according to the magnitude of activation [both Hb species reach statistical threshold ($P < 0.05$) = red, 1 Hb species reaching threshold ($P < 0.05$) = pink], deactivation [both Hb species reach statistical threshold ($P < 0.05$) = light blue, 1 Hb species reaching threshold ($P < 0.05$) = dark blue], or an absence of significant cortical hemodynamic change (white circles). B, Charts summarize group-averaged statistical analysis of HbO_2 and HHb and presented in the form of series of activation/deactivation matrices. Each plot represents an experience group (left column = novices, middle column = residents, right column = attendings) and the *primed* conditions either video review (i) or decision-making episodes (ii). Twenty-two channels are highlighted (black circles) and color coded to according to the magnitude of activation [both Hb species reach statistical threshold ($P < 0.05$) = red, 1 Hb species reaching threshold ($P < 0.05$) = pink], deactivation [both Hb species reach statistical threshold ($P < 0.05$) = light blue, 1 Hb species reaching threshold ($P < 0.05$) = dark blue], or an absence of significant cortical hemodynamic change (white circles).

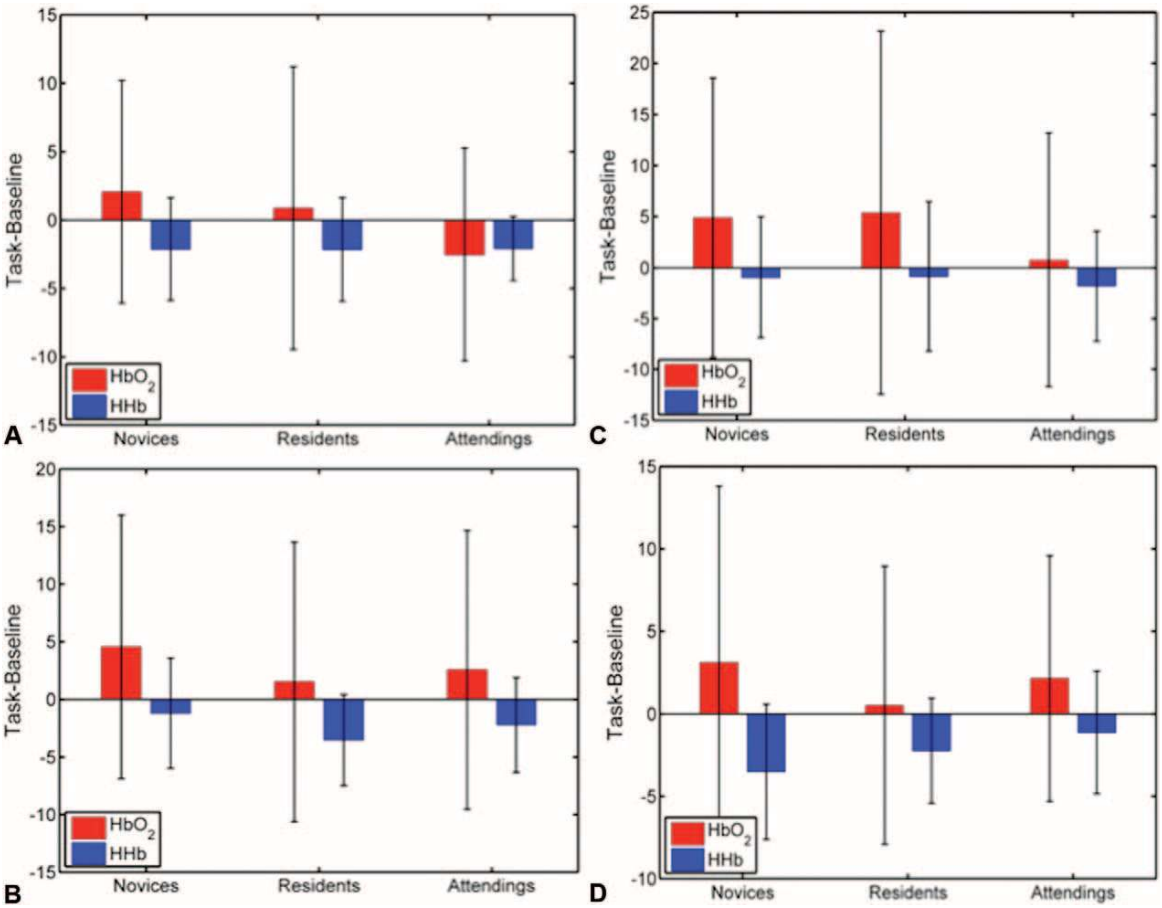


FIGURE 4. Bar charts illustrating between-group differences in mean ΔHbO_2 (red bars) and ΔHHb (blue bars) for certain right dorsolateral prefrontal channels (A, channel 22; B, channel 5) and left dorsolateral prefrontal channels (C, channel 1; D, channel 10).

AQ6 Table 1 shows comparisons between operators in ΔHb data during DM stimuli for bilateral DLPFC channels. DM-associated changes in cortical HbO_2 and HHb were substantially greater among novices vs operators with prior laparoscopic cholecystectomy experience. As shown in Figure 4, trends toward significantly greater activation responses in novices vs residents and attendings were observed in multiple bilateral DLPFC channels (ΔHbO_2: right DLPFC channel 22, ΔHHb: right DLPFC channels 5 and 13, and left DLPFC channel 10).

Primed Decisions

As shown in the averaged Hb time course curves (Supplementary Figure 2, http://links.lww.com/SLA/A972), in general, PFC responses during operative DM were less apparent in the primed vs the unprimed condition. Indeed, as shown in the matrices Figure 3 (panel b) regardless of expertise, priming did not lead to statistically significant activation either during video review or during DM stimuli. Rather during video review, an inverse relationship was identified between deactivation trends and operator expertise (deactivated channel trends: novices = 1/22, residents = 4/22, and attendings = 5/22). During DM trials, bilateral DLPFC activation trends were identified in novices and residents, whereas no significant cortical hemodynamic change was apparent among attendings.

Table 2 shows within-group GLMM results including the model's coefficients for the effect of the fixed factor (priming), which reveal the direction and significance of the effects. Overall, the priming effect was observed only for HbO_2 in novices—the significant negative coefficient implies that the priming reduced ΔHbO_2 across the DLPFC. However, a between-group GLMM model did not demonstrate an expertise × priming interaction effect [ΔHbO_2: $F(2786) = 0.56$, $P = .569$; ΔHHb: $F(2786) = 0.04$, $P = .957$].

DISCUSSION

In this study, expertise-related differences in intraoperative DM performance, consistency, and confidence have been investigated, and DM strategies have been exposed using functional neuroimaging. As hypothesized, expert DM was characterized by superior quality decisions, greater confidence in DM, and a willingness to challenge apparent decisions made by another operator. Furthermore, novice DM in the face of uncertainty (ie, absence of the behavioral cue or prime) was manifested as greater dorsolateral, ventrolateral, and medial PFC activations, suggesting a need for greater attention, concentration, and mental effort during DM. The results of within-group analysis suggest that the introduction of a behavioral trigger that revealed the operator's next operative decision prompted attenuation of prefrontal activation among novices. This notwithstanding, upon between-group analysis no such expertise × priming interaction effect was observed, most likely due to the relatively small numbers available for formal analysis.

Traditional pyramidal models of learning suggest that in the process of skills acquisition the learner transcends discrete phases associated with different mental processes.[30] Applying this model to skills in operative DM, progressive improvement is associated with transition from a novice phase that relies on a rigid adherence to taught "rules" or "goals" (goal-orientated DM) to an expert intuitive mode that relies on implicit knowledge and experience (habitual DM). Moreover, according to the work of Ericsson,[31] expertise in operative DM likely arises as a result of "deliberate practice" in which tasks are deconstructed and trained through formative feedback.[31] Similarly, emerging evidence indicates neural interactions occur in the transition from goal-directed to habitual DM.[32] Transition from goal-orientated to habitual DM is likely to take place during the acquisition of expertise in surgical DM. This is because habits require extensive experience including schedules of reinforcement involving actions and outcomes, indicating that behavior must be initially goal-directed before gradually becoming habitual over the course of experience.

Therefore, the observed increase in confidence and quality of DM among expert laparoscopists likely reflects years of repeated exposure to similar operative scenes and reflection regarding the outcomes of their own DM, and observation of resident DM. Habitual DM represents stimulus-response associations learned through repeated practice and rewards in a stable environment.[33] Habits are implemented in the subcortical structures—the dorsolateral striatum and dopamine neurons into this area, arriving from substantia nigra and the ventral tegmental area, are important for learning the value of habitual actions, and stimulus-response representations can also be encoded in corticothalamic loops and the infralimbic (medial) prefrontal cortex.[32] Hence, the relative DLPFC and MPFC redundancy during expert DM reflects the establishment of patterns of habitual DM, which is stable and repetitive with similar cues, actions, and rewards.

Conversely, the observed prefrontal activation response among novices suggests a goal-directed intraoperative DM approach. Goal-directed DM is implemented in different parts of the frontal lobe, concentrating on the anterior cingulate and orbitofrontal cortex, and also subsuming mechanisms localized in hippocampus and dorsomedial striatum.[18] Goal-directed decisions and actions are implemented predominantly in networks that mediate declarative expectations of future outcomes and conscious planning.[34,35] Effortful decisions depending on working memory and those that involve reasoning cause recruitment of the DLPFC[2,21,41] and the anterior cingulate cortex.[37,42,43] Decisions requiring cross-reference to the decision maker's value system, incorporation of long-term or contextual information and decisions made under uncertainty are known to burden the DLPFC.[20,38,45–47] Finally, goal-directed DM specifically involves the anterior cingulate cortex during highly ambiguous situations in which the decision maker perceives several conflicting options and a high likelihood of error,[37,38] which also may explain the relative PFC redundancy among novices during primed intraoperative DM.

InsertRef
39 here

It is interesting to note that when faced with an apparent decision made by another operator (ie, during surgical cues/behavioral primes), novices infrequently challenge the decision, possibly considering it to be the correct next operative move. Although subjects were not informed as to the operator's identity, novices may have assumed that operator was an expert attending. We speculate that in the minds of novices, this incorrectly reduces uncertainty and ambiguity and prompts them to accept the observed decision. This acceptance seems to manifest as a comparative prefrontal disengagement and lack of attention and concentration that was previously required for intraoperative DM under greater uncertainty, that is when what to do next was not obvious. In contrast, expert surgeons with greater experience and improved confidence more frequently challenge operative decisions that they perceive to be incorrect. This is unsurprising considering that in daily practice senior surgeons are required to routinely challenge the operative DM of more junior surgeons in training. Expert surgeons primed with the salient cues (ie, the behavioral prime in this case the next operative move) during familiar operative scenes automatically make the associated decision without further thought, hence the lack of activation in goal-directed decision regions.

In our view there is tremendous potential to use the findings of this experiment toward improvements in training and performance, as shown in Figure 5. There is increasing interest in mentoring and coaching to improve technical and cognitive skills such as judgment and DM,[48,49] including the potential of procedural videos to be used for safe and timely coaching.[48,49] Specifically, the current repository of operative videos coupled with recorded expert decisions can now be used to better train and assess residents in operative DM. Residents can now be subjected to these operative scenarios and their judgment compared and contrasted with the operative decisions of the expert panel. Script concordance enables the allocation of points based on the degree to which residents' DM aligns with those of experts, and proficiency benchmarks for DM assessment can now be established. Decision confidence, consistency, and the frequency with which residents' challenge decision deemed incorrect by experts can also now be incorporated in residency assessments. Moreover, it is feasible to design debriefing sessions to enable mentors to feedback to residents regarding the quality of their operative DM and coach them as to what experts chose to do when faced with similar anatomical scenarios. It is envisaged that as this field develops further, more challenging operative DM scenarios can be developed, acting as a series of decision "hurdles" for residents to overcome to support independent practice, with the aim of minimizing costs and morbidity of operative errors.

Fascinatingly, the current analysis suggests that it may be possible to derive proficiency benchmarks in operative DM based on the intensity of brain responses to simulated laparoscopic surgery. Specifically, intense DLPFC and VMPFC responses during unprimed decisions, and "inappropriate" PFC disengagement during primed decisions, seem to define the brain responses of novice operators. Similarly, the magnitude of brain responses may help expose instances when trainees are excessively ruminating and hence unsure of the next operative move (ie, excessive prefrontal changes). However, to capitalize on the benefits of functional imaging, neuroimaging technology must become more discrete and the analysis algorithms more automated to provide trainers with intelligible data regarding levels of resident attention and concentration in a similar fashion to metrics provided by virtual reality simulators. Portable, wearable, and wireless functional near-infrared spectroscopy systems are already in development and are set to become more affordable with less obtrusive headgear that can be discretely worn under the surgical hat. Our group and others are working on machine learning algorithms that can decode operator brain states online and that longer term could support implementation in residency programs.

Finally, mentoring, coaching, and cognitive biofeedback training that has already been shown to improve microsurgical skills[50] are interventions that may facilitate improved operative DM and increased decision confidence. Critically, by capitalizing on the current findings these interventions can now be tested to see if they result in more rapid attenuation of prefrontal brain responses among residents such that they align more closely with brain responses of

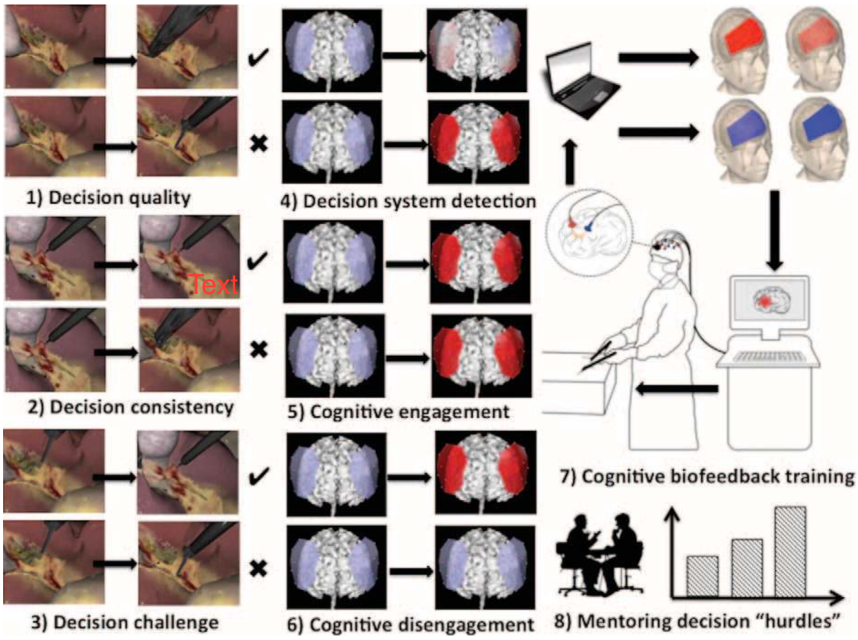


FIGURE 5. Schematic illustration summarizing short-term translation and long-term clinical impact, as follows: A, *Assessment of decision quality*—the validated set of operative videos and matched expert panel responses can be used to assess decision quality, evaluating the degree of alignment (upper panel—clip duct) or misalignment (lower panel—cut duct) between resident and expert surgical decision making; B, *Decision consistency*—residents' operative decision consistency can be assessed across similar but temporally spaced anatomical scenarios to determine the degree of consistency (upper panel) or inconsistency (lower panel) in operative decision making; C, *Decision challenge*—simulations that deliberately depict poor operative decisions determine whether residents are willing to "challenge" (upper panel) or simply "accept" erroneous decisions (lower panel); D, *Assessment of decision system*—the spatial distribution and intensity of brain activation provide insights into the decision system operators use, making it possible to detect shifts from the "goal-orientated" system of the novices (lower panel) to the "recognition primed systems" of experts (upper panel); E, *Cognitive engagement*—neuroimaging enables assessment of levels of cognitive engagement which are known to be important in formulating early decision outcome relationships and enables inappropriate *Cognitive disengagement* (F) to be detected. Finally, in the future with online analysis it may be possible to display maps of brain engagement/disengagement to the operator or trainer to enable "Cognitive Biofeedback" (G) designed to improve decision quality by augmenting attention and concentration. *Mentoring* (H) and progressive decision "hurdles" may improve resident readiness for independent practice in the operating room.

experts. Most importantly, unlike studies that raise the importance of assessing operator attention,[12] describe operative decision theory,[5] and generate qualitative cognitive taxonomy,[9] the current study objectively quantifies brain activation, demonstrates that the magnitude of executive control is related to surgical expertise in DM, and is timely when framed against the recent sea change from assessment solely of technical skills toward innovative approaches to assess attention, perception, and judgment in surgery. AQ7

In summary, attendings' DM is characterized by greater confidence, improved alignment with an expert reference panel, and reduced reliance on the prefrontal lobe suggesting mature habitual responses. Prefrontal excitation observed in novices implies that the transition from trainee to expert is coupled with a switch from goal-orientated to recognition-based DM.

LIMITATIONS

A number of limitations of this study should be acknowledged. Current OT techniques have limited depth penetration, the temporal resolution is inferior to electroencephalography (ie, latency from contemplating operative decision to detecting a response), and the spatial resolution is inferior to functional magnetic resonance imaging. However, OT enables an operator's brain function to be interrogated during a realistic simulation of operative DM, provides objective hemodynamic data regarding which brain areas are recruited, and is more reliable than subjective responses. The nature of the experimental paradigm and time required for each subject (eg, approximately 1 h per subject for training, OT probe placement, task familiarization, and experiment) limited the recruitment of attendings. Although script concordance is a valid measure of agreement with panel consensus, it does not necessarily follow that the operative decisions made by attendings or indeed the expert panel were all "correct." Indeed, the concept of a single correct next operative decision is challenging to validate, and it is more likely that for a given scenario one of several options is safe. This notwithstanding, the aim was to explore the internal cognitive process and cortical responses associated with operative DM and these are not influenced by the specific decision. Put simply, the study primarily sought to address how a decision was arrived at, as opposed to whether the decision was correct or not. It should be acknowledged that the time set aside for DM after video review is artificial, and the internal processing regarding operative decisions is likely to be made continually online. However, the experiment was designed to enable us to isolate DM-associated cortical activations, which would not have been feasible in a less controlled experiment. Finally, we accept that given novices felt less stressed after the experiment, stress-induced changes in hemodynamics may have contributed to our results.

Uncited references

AQ8 [36,39,40,44].

REFERENCES

1. Endsley M. Toward a theory of situation awareness indynamic systems. *Hum Factors*. 1995;37:32–64.
2. Leff DR, Orihuela-Espina F, Athanasiou T, et al. ‘‘Circadian cortical compensation’’: a longitudinal study of brain function during technical and cognitive skills in acutely sleep-deprived surgical residents. *Ann Surg*. 2010;252:1082–1090.
3. Graafland M, Schraagen JM, Boermeester MA, et al. Training situational awareness to reduce surgical errors in the operating room. *Br J Surg*. 2015; 102:16–23.
4. Pauley K, Flin R, Yule S, et al. Surgeons’ intraoperative decision making and risk management. *Am J Surg*. 2011;202:375–381.
5. Flin R, Youngson G, Yule S. How do surgeons make intraoperative decisions? *Qual Saf Health Care*. 2007;16:235–239.
6. Rangel A, Camerer C, Montague PR. A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci*. 2008;9:545–556.
7. Dolan RJ, Dayan P. Goals and habits in the brain. *Neuron*. 2013;80:312–325.
8. Way LW, Stewart L, Gantert W, et al. Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. *Ann Surg*. 2003;237:460–469.
9. Madani A, Watanabe Y, Feldman LS, et al. Expert intraoperative judgment and decision-making: defining the cognitive competencies for safe laparoscopic cholecystectomy. *J Am Coll Surg*. 2015;221:931–940e8.
10. Alkhaffaf B, Decadt B. 15 years of litigation following laparoscopic cholecystectomy in England. *Ann Surg*. 2010;251:682–685.
11. Carroll BJ, Birth M, Phillips EH. Common bile duct injuries during laparoscopic cholecystectomy that result in litigation. *Surg Endosc*. 1998;12:310–313.
12. Gallagher AG, Satava RM, O’Sullivan GC. Attentional capacity: an essential aspect of surgeon performance. *Ann Surg*. 2015;261:e60–e61.
13. Zilbert NR, Murnaghan ML, Gallinger S, et al. Taking a chance or playing it safe: reframing risk assessment within the surgeon’s comfort zone. *Ann Surg*. 2015;262:253–259.
14. Pauley K, Flin R, Azuara-Blanco A. Intra-operative decision making by ophthalmic surgeons. *Br J Ophthalmol*. 2013;97:1303–1307.
15. Cristancho SM, Vanstone M, Lingard L, et al. When surgeons face intraoperative challenges: a naturalistic model of surgical decision making. *Am J Surg*. 2013;205:156–162.
16. Seymour B, Singer T, Dolan R. The neurobiology of punishment. *Nat Rev Neurosci*. 2007;8:300–311.
17. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*. 2005;8:1704–1711.
18. Glimcher PW. Choice: towards a standard back-pocket model. *Neuroeconomics: Decision making and the brain* 2009:503–521. AQ9
19. Rolls ET. *Emotion Explained.* Oxford: Oxford University Press; 2005.
20. Dixon ML, Christoff K. The lateral prefrontal cortex and complex value-based learning and decision making. *Neurosci Biobehav Rev*. 2014;45:9–18.
21. Raine A, Yang Y. Neural foundations to moral reasoning and antisocial behavior. *Soc Cogn Affect Neurosci*. 2006;1:203–213.
22. Vlaev I, Dolan P. Action change theory: a reinforcement learning perspective on behaviour change. *Gen Pyschol*. 2015;19:69–95.
23. Orihuela-Espina F, Leff DR, James DR, et al. Quality control and assurance in functional near infrared spectroscopy (fNIRS) experimentation. *Phys Med Biol*. 2010;55:3701–3724.
24. Jobsis FF. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science*. 1977;198:1264–1267.
25. Roy CS, Sherrington CS. On the regulation of blood-supply to the brain. *J Physiol*. 1890;26:86–108.
26. Cope M, Delpy DT, Reynolds EO, et al. Methods of quantitating cerebral near infrared spectroscopy data. *Adv Exp Med Biol*. 1988;222:183–189.
27. Jasper HH. The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol Suppl*. 1958;10:371–375.
28. Tachtsidis I, Leung TS, Devoto L, et al. Measurement of frontal lobe functional activation and related systemic effects: a near-infrared spectroscopy investigation. *Adv Exp Med Biol*. 2008;614:397–403.
29. Charlin B, Roy L, Brailovsky C, et al. The Script Concordance test: a tool to assess the reflective clinician. *Teach Learn Med*. 2000;12:189–195.
30. Dreyfus S DH. A five stage model of the mental activities involved in directed skill acquisition. California University Berkeley Operations Research Center [monograph on the Internet]. February 1980; 1–12. Available from: Defense Technical Information Center, Belvoir, VA. Accessed October 26, 2015.
31. Ericsson KA, Nandagopal K, Roring RW. Toward a science of exceptional achievement: attaining superior performance through deliberate practice. *Ann N Y Acad Sci*. 2009;1172:199–217.
32. Graybiel AM. Habits, rituals, and the evaluative brain. *Annu Rev Neurosci*. 2008;31:359–387.
33. de Wit S, Watson P, Harsay HA, et al. Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *J Neurosci*. 2012;32:12066–12075.
34. Berridge KC, Kringelbach ML. Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology*. 2008;199:457–480.
35. Dolan Ray J, Dayan P. Goals and habits in the brain. *Neuron*. 2013;80:312–325.
36. Volz KG, Schubotz RI, von Cramon DY. Decision-making and the frontal lobes. *Curr Opin Neurol*. 2006;19:401–406.
37. Rosenbloom MH, Schmahmann JD, Price BH. The functional neuroanatomy of decision-making. *J Neuropsychiatry Clin Neurosci*. 2012;24:266–277.
38. Krain AL, Wilson AM, Arbuckle R, et al. Distinct neural mechanisms of risk and ambiguity: a meta-analysis of decision-making. *Neuroimage*. 2006;32:477–484.
39. Sescousse G, Redoute J, Dreher JC. The architecture of reward value coding in the human orbitofrontal cortex. *J Neurosci*. 2010;30:13095–13104.
40. Wallis JD. Orbitofrontal cortex and its contribution to decision-making. *Annu Rev Neurosci*. 2007;30:31–56.
41. Moll J, de Oliveira-Souza R. Moral judgments, emotions and the utilitarian brain. *Trends Cogn Sci*. 2007;11:319–321.
42. Mulert C, Seifert C, Leicht G, et al. Single-trial coupling of EEG and fMRI reveals the involvement of early anterior cingulate cortex activation in effortful decision making. *Neuroimage*. 2008;42:158–168.
43. Esposito F, Mulert C, Goebel R. Combined distributed source and single-trial EEG–fMRI modeling: application to effortful decision making processes. *Neuroimage*. 2009;47:112–121.
44. Gläscher J, Adolphs R, Damasio H, et al. Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proc Natl Acad Sci*. 2012;109:14681–14686.
45. Causse M, Peran P, Dehais F, et al. Affective decision making under uncertainty during a plausible aviation task: an fMRI study. *Neuroimage*. 2013;71:19–29.
46. Catena A, Perales JC, Megias A, et al. The brain network of expectancy and uncertainty processing. *PLoS One*. 2012;7:e40252.
47. Huettel SA, Song AW, McCarthy G. Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *J Neurosci*. 2005;25:3304–3311.
48. Bonrath EM, Dedy NJ, Gordon LE, et al. Comprehensive surgical coaching enhances surgical skill in the operating room: a randomized controlled trial. *Ann Surg*. 2015;262:205–212.
49. Greenberg CC, Ghousseini HN, Pavuluri Quamme SR, et al. Surgical coaching for individual performance improvement. *Ann Surg*. 2015;261:32–34.
50. Ros T, Moseley MJ, Bloom PA, et al. Optimizing microsurgical skills with EEG neurofeedback. *BMC Neurosci*. 2009;10:87.

ANNSURG
Manuscript No. 15-01288

Annals of Surgery

Dear Author,

During the preparation of your manuscript for typesetting, some queries have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin as neatly as possible or compile them as a separate list. This form should then be returned with your marked proof/list of corrections to the Production Editor.

QUERIES: to be answered by AUTHOR

QUERY NO.	QUERY DETAILS	RESPONSE
<AQ1>	Please confirm whether surnames/family names (red) have been identified correctly in the author byline.	AQ1 = surnames are correct
<AQ2>	Please check the affiliation for correctness.	AQ2=affiliations correct
<AQ3>	Please check correspondences details for appropriateness.	AQ3=reprints address correct
<AQ4>	Please check if the edit made to the sentence ‘‘To investigate differences in the quality, confidence, and consistency of intraoperative surgical decision making (DM) and decision systems that operators use using functional neuroimaging’’ retains the intended meaning.	AQ4= please change to read: To investigate differences in the quality, confidence, and consistency of intraoperative surgical decision making (DM) and using functional neuroimaging expose decision systems that operators use.
<AQ5>	Please verify the accuracy of the conflicts of interest statements.	AQ5=The authors have no conflict of interest. We would like to acknowledge the work was funded through the support of the Imperial NIHR Biomedical Research Centre and Imperial Cancer Research UK Centre
<AQ6>	Tables 1 and 2 are cited in the text, but not provided. Please check and confirm.	AQ6=Table 1 and 2 provided and re-attached through the online portal
<AQ7>	Please check the sentence ‘‘Most importantly, unlike studies that raise the importance . . . to assess attention, perception, and judgment in surgery’’ for clarity.	AQ7= please leave sentence as is
<AQ8>	Refs. [36, 39, 40, 44] are not cited in the text. Please check and confirm,	AQ8 = please delete refs 36, 40, and 44 which following revision are no longer required
<AQ9>	Please provide complete bibliographic details for the Refs. [18, 30].	
<AQ10>	Please check the changes made in all the figure captions for correctness.	

AQ9= Please replace ref 18 with the following:
Glimcher, PW. Choice: towards a standard back-pocket model. In: Glimcher, PW.; Camerer, CF.; Fehr, E.; Poldrack, RA., editors. Neuroeconomics: decision making and the brain. Amsterdam: Elsevier; 2008. p. 501-519.
Please replace ref 30 with this citation:
Dreyfus SE. The Five-Stage Model of Adult Skill Acquisition. Bulletin of Science Technology & Society 2004; 24:177-181.

AQ10=Changes marked up in the Figure legends as requested