

# Impact of simulated three-dimensional perception on precision of depth judgements, technical performance and perceived workload in laparoscopy

S. Sakata<sup>1,3,4</sup>, P. M. Grove<sup>2</sup>, A. Hill<sup>2,4</sup>, M. O. Watson<sup>1,2,4</sup> and A. R. L. Stevenson<sup>1,3</sup>

Schools of <sup>1</sup>Medicine and <sup>2</sup>Psychology, The University of Queensland, and <sup>3</sup>Department of Colon and Rectal Surgery, Royal Brisbane and Women's Hospital, and <sup>4</sup>Clinical Skills Development Service, Metro North Hospital and Health Service, Brisbane, Queensland, Australia

Correspondence to: Dr S. Sakata, School of Medicine, The University of Queensland, Mayne Medical Building, Herston Road, Herston, Queensland 4006, Australia (e-mail: shin.sakata@gmail.com)

**Background:** This study compared precision of depth judgements, technical performance and workload using two-dimensional (2D) and three-dimensional (3D) laparoscopic displays across different viewing distances. It also compared the accuracy of 3D displays with natural viewing, along with the relationship between stereoacuity and 3D laparoscopic performance.

**Methods:** A counterbalanced within-subjects design with random assignment to testing sequences was used. The system could display 2D or 3D images with the same set-up. A Howard–Dolman apparatus assessed precision of depth judgements, and three laparoscopic tasks (peg transfer, navigation in space and suturing) assessed performance (time to completion). Participants completed tasks in all combinations of two viewing modes (2D, 3D) and two viewing distances (1 m, 3 m). Other measures administered included the National Aeronautics and Space Administration Task Load Index (perceived workload) and the Randot<sup>®</sup> Stereotest (stereoacuity).

**Results:** Depth judgements were 6.2 times as precise at 1 m and 3.0 times as precise at 3 m using 3D versus 2D displays ( $P < 0.001$ ). Participants performed all laparoscopic tasks faster in 3D at both 1 and 3 m ( $P < 0.001$ ), with mean completion times up to 64 per cent shorter for 3D versus 2D displays. Workload was lower for 3D displays (up to 34 per cent) than for 2D displays at both viewing distances ( $P < 0.001$ ). Greater viewing distance inhibited performance for two laparoscopic tasks, and increased perceived workload for all three ( $P < 0.001$ ). Higher stereoacuity was associated with shorter completion times for the navigating in space task performed in 3D at 1 m ( $r = -0.40$ ,  $P = 0.001$ ).

**Conclusion:** 3D displays offer large improvements over 2D displays in precision of depth judgements, technical performance and perceived workload.

Paper accepted 7 February 2017

Published online 20 April 2017 in Wiley Online Library (www.bjs.co.uk). DOI: 10.1002/bjs.10528

## Introduction

Stereopsis is the perception of depth that arises from comparison of slight differences, called disparities, in the images that project to two laterally separated eyes<sup>1</sup>. The three-dimensional (3D) systems used in minimally invasive surgery capture separate left- and right-eye images with a dual-channel laparoscope, and simulate the binocular images that would result if the viewer were positioned at the tip of the laparoscope<sup>2,3</sup>. In modern 3D laparoscopy, images are viewed via passive polarization in which the viewer wears lightweight glasses that polarize horizontal rows of pixels on the display, with alternate pixel rows

corresponding to the right- and left-eye images. Conventional two-dimensional (2D) displays do not provide stereopsis, and laparoscopic surgeons must rely on indirect visual cues such as shadows, textures and relative colour differences to extract depth information. These indirect cues do not provide a compelling or immediate perception of 3D volume and may not be appreciated by inexperienced viewers<sup>1</sup>.

Current understanding of the relative advantages of 3D over 2D laparoscopy is based primarily on research with confounded study designs<sup>2</sup>. Over the past 20 years, studies comparing these systems have not considered the impact of an individual surgeon's stereoacuity, viewing distance

and cross-talk on the precision of depth judgements gained from laparoscopic 3D displays<sup>2</sup>. The failure to account for these basic perceptual factors may have contributed to the inconsistent and conflicting evidence addressing the benefit of 3D over 2D laparoscopy in terms of technical performance and patient outcomes<sup>2</sup>.

An individual's stereoacuity is defined as the smallest depth interval signalled by binocular disparity that can be detected reliably<sup>1</sup>. Sensitivity to binocular disparity information varies substantially in the general population, with two-thirds of individuals having good to excellent stereoacuity (so-called 'haves') and the remainder having moderate to poor stereoscopic perception (so-called 'have-nots')<sup>4</sup>. A recent cross-sectional study<sup>5</sup> of surgeons suggests that 10 per cent may be 'stereoblind'. Binocular disparity information is also affected by viewing distance. In direct viewing, the binocular disparity arising from a fixed depth interval is inversely proportional to the viewing distance. The impression of depth generated by a given disparity in a 3D display changes with viewing distance, which can result in distortions of depth relationships in the display. As an individual's ability to perceive a given distance between two objects is affected by viewing distance<sup>1</sup>, studies comparing 3D with 2D laparoscopic displays must also take account of the viewing distance of their participants.

Cross-talk results from viewing a 3D display from sub-optimal viewing locations, such that the orientation of the polarized glasses is not matched to the display monitor. Consequently, each eye sees a mixture of the image intended for that eye and parts of the image intended for the other eye, referred to as 'ghosting'<sup>6</sup>. Cross-talk impairs stereovision, and in severe cases results in headache, disorientation and fatigue<sup>7</sup>. There are common working positions in the operating theatre that may be suboptimal for surgical team members sharing a 3D display, where even short periods of viewing can yield high levels of ghosting and discomfort<sup>6</sup>.

Given the conflicting evidence as to whether the perception of simulated 3D volume translates directly into significant improvements in surgical performance, it was hypothesized that these inconclusive findings may be due to studies not accounting for individual differences in stereoacuity or situational variables such as viewing distance and/or cross-talk when comparing 3D with 2D display systems<sup>2</sup>.

The primary aim of this study was to compare precision of depth judgements, technical performance and perceived workload using 2D and 3D laparoscopic displays. As the viewing distance between the laparoscopic surgeon and the 3D display may vary in clinical practice, participants were assessed on both viewing modes at two viewing distances

to increase the generality of the results, and their viewing positions were optimized to minimize cross-talk during 3D laparoscopy. Secondary aims of this study were to assess the accuracy of a modern 3D laparoscopic display compared with natural binocular viewing, to determine which laparoscopic technical tasks require precise depth judgements, and to examine the relationship between individual stereoacuity and 3D laparoscopic performance.

## Methods

A counterbalanced within-subjects design was used in which participants were assigned randomly to one of 32 unique testing sequences. Some 64 volunteers with a mean age of 27.5 years took part in the study. Participants were junior doctors from the University of Queensland (UQ) who were members of the UQ Surgical Interest Group. They were the first 64 respondents to an advertisement placed on the UQ Surgical Interest Group webpage. To minimize floor effects, all participants had previously completed a laparoscopic skills course taught by a surgeon instructor at the Clinical Skills Development Service, Brisbane, Australia.

### Measurement of individual stereoacuity using the Randot<sup>®</sup> Stereotest

The Randot<sup>®</sup> Stereotest (Stereo Optical, Chicago, Illinois, USA) presents ten sets of three circles. In each set, the target circle has crossed disparity (appears to be closer to the viewer than the other two circles) when viewed through testing glasses containing cross-polarized filters. Across sets, the target circles have sequentially decreasing disparities of 400 to 20 seconds of arc, and the participant's task is to identify which circle appears closer than the other two. Threshold stereoacuity is determined by the last set in which the participant is able correctly to identify the target circle<sup>8</sup>.

### Screening participants for contrast impairment

Contrast sensitivity is the ability of the visual system to detect differences in luminance between an object and its background<sup>9</sup>. The Pelli–Robson contrast sensitivity acuity chart (Clement Clarke International, Harlow, UK) is the most widely used contrast sensitivity measure and uses large letters as targets<sup>9</sup>. For each consecutive group of three letters, contrast decreases from left to right and from top to bottom of the chart. The lowest contrast at which two or three of the letters in a group can be read determines the participant's log contrast sensitivity score.

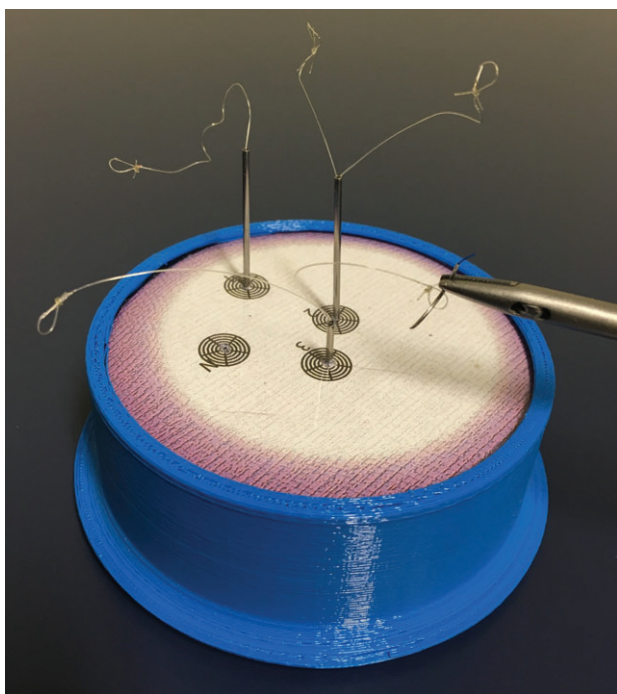


Fig. 1 Apparatus for the navigating in space laparoscopic task

A score of 2.0 indicates optimal contrast sensitivity (100 per cent), whereas a score below 1.5 suggests a sensitivity impairment.

### Screening participants for visual acuity impairment

The logarithm of the minimum angle of resolution or log-MAR visual acuity chart (National Vision Research Institute, Melbourne, Australia) is the accepted standard for measuring visual acuity<sup>10,11</sup>. The logMAR uses rows of letters that become progressively smaller and has a constant 0.10 log unit difference between each row. Raw scores are converted into a visual acuity score (VAS) where a score of 100 corresponds to 6/6 vision and higher values correspond to better visual acuities<sup>11</sup>. For screening purposes, VAS values below 100 were considered indicative of visual acuity impairment.

### Minimizing cross-talk by optimizing viewing position

Cross-talk or 'ghosting' is produced by incorrect orientation of the polarized glasses relative to the display monitor, resulting in each eye seeing a mixture of the image intended for that eye and parts of the image intended for the other eye. To minimize cross-talk, each participant was tested in the optimal viewing position in which the centre of the

plane of the display was adjusted to be perpendicular to the viewer's line of sight<sup>7</sup>.

### Comparing 3D and 2D images using the same system to prevent confounds

The laparoscopic system used was capable of displaying 2D or 3D images with the same set-up. Target images were captured by the Olympus Endoeye Flex 3D laparoscope, with left and right images relayed to individual Olympus CV-190 processors and integrated by an Olympus 3DV-190 visualization unit (Olympus Corporation, Tokyo, Japan). The resulting images were displayed on a compatible Sony LMD-2451MT LCD HDTV monitor (Sony Corporation, Tokyo, Japan) (dimensions: 60.3 × 38.7 cm), which incorporates a high-resolution LCD panel (1900 × 1200 pixels) and is intended for medical use only. In the 3D conditions, the display was set for 3D presentation and the participants wore passive polarizing glasses. In the 2D conditions, the display was set to 2D and participants viewed it without polarizing glasses.

### Varying the viewing distance systematically

Perception of simulated depth is affected by viewing distance and this, in turn, may affect technical performance in 3D laparoscopy<sup>1</sup>. Two viewing distance conditions were therefore used to approximate common viewing distances in laparoscopic surgery (1 and 3 m). Viewing distance was measured as the straight distance from the back of the participant's heels to the mid-point of a horizontal line on the floor parallel to the plane of the display monitor.

### Testing precision of depth judgements

The Howard–Dolman apparatus is used to test the precision of an individual's judgements of the relative depth of objects in space with minimization of indirect depth cues. Participants were required to align two vertical sutures (one blue, one undyed) as closely as possible in depth, so that both sutures appeared to lie side-by-side in the same frontoparallel plane. The position of the undyed suture was fixed at 10 cm from the viewing aperture and participants could control the position of the blue suture by pulling on strings attached to a track mechanism. Precision was expressed as the just noticeable difference (JND), a basic parameter of visual performance, synonymous with human variable error<sup>12</sup>. In the context of this task, a participant's JND represented the smallest depth interval between the

fixed and adjustable sutures that they could discern reliably. Before this study, all participants were naive to the Howard–Dolman apparatus.

To test depth perception using laparoscopic 2D and 3D images, the laparoscope was mounted so that its tip was directed towards the aperture of the apparatus and perpendicular to the sutures. The distance from the laparoscope to the apparatus was set so that the image of the aperture on the monitor had the same dimensions as the actual aperture. While completing the task using the monitor, participants were positioned so that their line of sight was perpendicular to its centre. Participants were tested at 1 and 3 m.

### Accuracy of depth judgements in 3D laparoscopy versus direct viewing

Accuracy, or constant error, is defined as the level of bias associated with a perceptual judgement, and is calculated by averaging the signed deviations from the standard stimulus<sup>12</sup>. For the laparoscope-mediated Howard–Dolman task described above, accuracy measurements would reveal any perceptual bias associated with depth judgements using the 3D laparoscopic display, the Howard–Dolman apparatus, or both. To compare accuracy between direct binocular depth perception and the 3D display, participants also completed the Howard–Dolman task while viewing the sutures directly through the aperture with both eyes (and holding their heads still). In relation to the signed mean (calculated across all participants' mean scores), a positive sign would indicate that, on average, the adjustable blue suture was set by the participant to be more distant than the fixed undyed suture, whereas a negative sign would indicate that, on average, the adjustable blue suture was set closer to the observer than the fixed undyed suture. Ideally, if both the 3D display and the Howard–Dolman apparatus were perfect, the constant error or bias associated with each would be zero.

### Testing laparoscopic performance with technical tasks

Peg transfer skill was tested using the standard Fundamentals of Laparoscopic Surgery task module. This task tests gross instrument control in manipulating objects in 3D space, under time pressure. The time taken to transfer 24 coloured plastic triangles on to the pegs of an empty pegboard was recorded using a stopwatch.

The navigating in space task used a custom apparatus developed for the study (*Fig. 1*). This task tests fine instrument control with a needle in 3D space, under time pressure. Participants were required to hold a laparoscopic

needle-holder with their dominant hand, and pass a curved needle fixed in its jaws through a 2-mm loop at the tip of a monofilament suture. The apparatus incorporated five such loops that participants were required to complete in a pre-determined sequence. The time taken for the needle to pass through all five loops was recorded using a stopwatch.

Laparoscopic suturing skill was tested using a standard Fundamentals of Laparoscopic Surgery suturing task module. This task tests fine instrument control with a needle and suture, as well as suturing technique in 3D space, under time pressure. Participants were required to handle one laparoscopic needle-holder with each hand, and manipulate a curved needle on a 12-cm length of 3/0 polypropylene suture to repair a defect in a rubber tube. A black dot on the edge of each defect indicated where the suture must enter and exit. The time taken for a complete and flat laparoscopic knot (2, then 1, and then 1 throw) was recorded with a stopwatch.

### Evaluating perceived workload during laparoscopic tasks

Workload is a term used in human factors science to describe the mental and physical costs incurred by an individual during work activity. The National Aeronautics and Space Administration Task Load Index (NASA TLX) questionnaire was initially developed to analyse the perceived workload of pilots with the aim of redesigning processes to reduce technical errors, and has become the recognized standard for workload assessment with extensive validation in different industries<sup>13,14</sup>. The NASA TLX questionnaire has also been used to measure the perceived workload of surgeons and physicians trialling new procedures and technologies designed to improve work efficiency and performance<sup>15–18</sup>.

### Procedure

The testing procedures were approved by the UQ Human Research Ethics Committee and the Programme Director of the UQ School of Medicine. Participants were tested individually.

At the beginning of the session, participants completed the Randot<sup>®</sup> Stereotest and their ocular dominance was assessed using the near-point test. Participants then completed the remaining tasks in an order determined by random assignment to one of 32 counterbalanced testing sequences (2 participants per sequence). The 32 sequences comprised a factorial combination of: 2 vision testing orders (monocular/binocular or binocular/monocular) × 2 superordinate task orders (technical tasks/Howard–Dolman tasks or Howard–Dolman

tasks/technical tasks) × 2 Howard–Dolman testing mode orders (natural viewing/laparoscopic or laparoscopic/natural viewing) × 2 viewing mode orders (applied to each individual technical task and Howard–Dolman testing mode: 2D/3D or 3D/2D) × 2 viewing distance orders (applied within each viewing mode block for each individual technical task and Howard–Dolman testing mode: 1 m/3 m or 3 m/1 m).

In each testing sequence, there were two separate rounds of visual acuity and contrast sensitivity testing: monocular (using the dominant eye, with the non-dominant eye patched) and binocular (using both eyes). To minimize the risk of chart memorization, these two rounds of vision testing were kept separate, with one performed before the experimental tasks and the other held at the end of the session.

When tested with the Howard–Dolman apparatus, each participant performed the task six times for each combination of Howard–Dolman testing mode (natural viewing, laparoscopic), viewing mode (3D, 2D) and viewing distance (1 m, 3 m). When tested in the natural viewing mode, 3D viewing comprised natural binocular viewing of the Howard–Dolman apparatus, and 2D viewing involved natural monocular viewing using only the dominant eye. However, data from the natural monocular viewing condition are not relevant to the present aims and are not reported here.

All participants performed the laparoscopic technical tasks in order of difficulty: peg transfer, navigating in space and suturing. Participants completed each task three times for each combination of viewing mode (3D, 2D) and distance (1 m, 3 m), and then completed a NASA TLX questionnaire before moving to the next combination or task. Participants were asked whether they had experienced double vision during 3D laparoscopic viewing.

### Statistical analysis

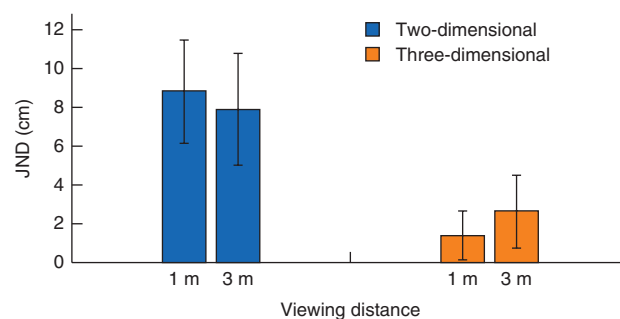
To assess the precision of depth judgements in the laparoscopy system, a separate JND value (expressed in centimetres) was calculated from each participant's Howard–Dolman task data for each combination of viewing mode (2D, 3D) and viewing distance (1 m, 3 m). Each JND value was derived by taking the standard deviation of the individual's settings (positive and negative deviations from the zero point where the sutures aligned perfectly) and multiplying by the constant 0.6745<sup>12,19,20</sup>.

Statistical analyses were conducted using SPSS<sup>®</sup> version 22 (IBM, Armonk, New York, USA) with  $\alpha$  set at 0.05. A repeated-measures ANOVA was conducted on the laparoscopic JND data to assess the effects of viewing mode (2D versus 3D) and viewing distance (1 m versus 3 m). For

**Table 1** Number of participants who obtained each possible threshold stereoacuity score on the Randot<sup>®</sup> Stereotest

Threshold stereoacuity (seconds of arc)	Randot <sup>®</sup> score	No. of participants (n = 64)
20	10*	31 (48)
40	9	8 (13)
50	8	11 (17)
60	7	1 (2)
80	6	4 (6)
100	5	6 (9)
140	4	2 (3)
400	2	1 (2)
800	1	0 (0)
Stereoblind	0	0 (0)

Values in parentheses are percentages. \*Best stereoacuity.



**Fig. 2** Precision of depth judgements by participants using the Howard–Dolman apparatus via the laparoscopic system. Values are mean (s.d.) just noticeable differences (JNDs), arranged according to viewing mode (two-dimensional versus three-dimensional) and viewing distance (1 m versus 3 m)

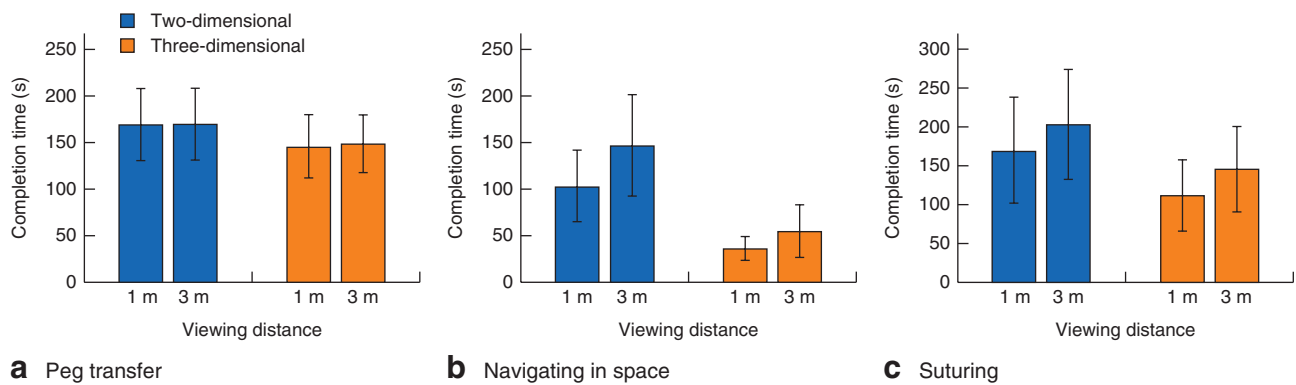
each laparoscopic task (peg transfer, navigating in space and laparoscopic suturing), similar repeated-measures ANOVAs were then conducted on the performance (time to completion) and workload data. Each significant viewing mode × viewing distance interaction was followed up with simple effects tests (paired *t* tests) comparing 2D versus 3D at each distance.

The effects of stereoacuity on laparoscopic performance were examined with a series of 12 Pearson correlation coefficients assessing the relationship between Randot scores (1–10) and time to completion for each of the three laparoscopic tasks under each combination of viewing mode (2D, 3D) and viewing distance (1 m, 3 m). To interpret these correlations,  $\alpha$  was adjusted using the Bonferroni correction for multiple comparisons ( $\alpha = 0.004$ ).

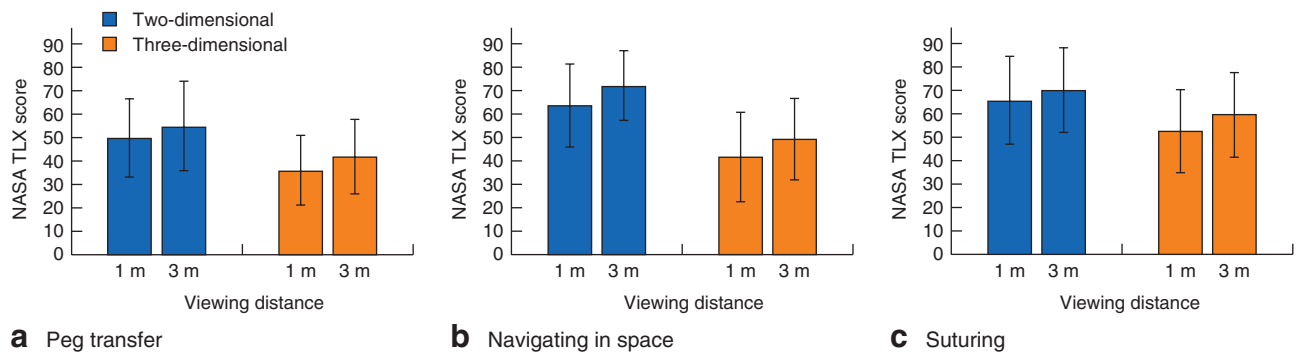
## Results

### Stereoacuity

The numbers of participants who obtained each possible threshold stereoacuity score on the Randot<sup>®</sup> Stereotest are



**Fig. 3** Laparoscopic performance (mean(s.d.) completion time) according to viewing mode (two-dimensional *versus* three-dimensional) and viewing distance (1 m *versus* 3 m) for three tasks: **a** peg transfer, **b** navigating in space and **c** suturing



**Fig. 4** Perceived workload (mean(s.d.) National Aeronautics and Space Administration Task Load Index (NASA TLX) score) according to viewing mode (two-dimensional *versus* three-dimensional) and viewing distance (1 m *versus* 3 m) for three tasks: **a** peg transfer, **b** navigating in space and **c** suturing

shown in *Table 1*. None of the 64 participants was found to be stereoblind.

### Eye dominance, contrast sensitivity, visual acuity and double vision

Thirty-one participants were left-eye dominant and 33 were right-eye dominant. Using the Pelli–Robson contrast sensitivity chart, no participant was found to have an impairment (score less than 1.5) in binocular contrast sensitivity (mean score 1.85) or dominant-eye monocular contrast sensitivity (mean score 1.78). Using the logMAR visual acuity chart, no participant had a VAS of less than 100 corresponding to visual acuity poorer than 6/6 vision (binocular VAS, mean 103.9; monocular VAS, mean 103.3). During laparoscopic 3D viewing, no participant reported experiencing double vision.

### Precision of depth judgements made via the laparoscopy system

Analysis of the laparoscopic precision data (*Fig. 2*) revealed a significant main effect of viewing mode

( $F(1,63)=447.59$ ,  $P<0.001$ ), qualified by a significant viewing mode  $\times$  viewing distance interaction ( $F(1,63)=32.85$ ,  $P<0.001$ ). At 3 m, the JND for depth judgements made via the laparoscopy system was 5.26 (95 per cent c.i. 4.49 to 6.02) cm smaller for 3D than for 2D images ( $t(63)=13.67$ ,  $P<0.001$ ). At 1 m, the advantage of 3D images was even greater: the JND was 7.38 (6.75 to 8.01) cm smaller for 3D *versus* 2D images ( $t(63)=23.42$ ,  $P<0.001$ ). In relative terms, depth judgements made using 3D images rather than 2D images were 3.0 times as precise at 3 m and 6.2 times as precise at 1 m (calculated using the values presented in *Fig. 2*, by dividing the precision value for 2D by the precision value for 3D for each viewing distance).

### Accuracy (perceptual bias)

Accuracy was the level of perceptual bias associated with depth judgements. When viewing the same stimulus (the Howard–Dolman apparatus), the levels of bias (mean deviation from zero) for direct binocular viewing and the 3D

**Table 2** Correlation coefficients assessing the relationship between Randot® stereoacuity scores and time to completion for each laparoscopic task under each combination of viewing mode and viewing distance

Laparoscopic task	Viewing at 1 m		Viewing at 3 m	
	2D	3D	2D	3D
Peg transfer				
<i>r</i>	-0.03	0.00	0.07	-0.13
<i>P</i>	0.793	1.000	0.609	0.308
Navigating in space				
<i>r</i>	-0.17	-0.40*	-0.25	-0.19
<i>P</i>	0.192	0.001	0.051	0.135
Laparoscopic suturing				
<i>r</i>	0.02	-0.08	-0.05	-0.04
<i>P</i>	0.856	0.557	0.678	0.765

2D, two-dimensional; 3D, three-dimensional. \*Significant at  $\alpha = 0.004$  ( $\alpha$ -adjusted using the Bonferroni correction for multiple comparisons).

display, respectively, were 0.1 *versus* -0.7 cm at 1 m, and 0.3 *versus* -0.8 cm at 3 m.

### Laparoscopic performance

#### Peg transfer task

ANOVA on the peg transfer task performance data indicated that participants completed the task significantly faster when viewing in 3D compared with 2D ( $F(1,63) = 70.65$ ,  $P < 0.001$ ) (Fig. 3a). In relative terms, completion times were 14 per cent shorter at 1 m and 13 per cent shorter at 3 m when viewing in 3D *versus* 2D (calculated using the values presented in Fig. 3a, by deducting 3D completion time from 2D completion time and then dividing by 2D completion time for each viewing distance).

#### Navigating in space

Significant main effects involved viewing mode ( $F(1,63) = 278.06$ ,  $P < 0.001$ ) and viewing distance ( $F(1,63) = 72.40$ ,  $P < 0.001$ ), qualified by a significant viewing mode  $\times$  viewing distance interaction ( $F(1,63) = 16.29$ ,  $P < 0.001$ ) (Fig. 3b). At 1 m, participants were 65.72 (95 per cent c.i. 56.22 to 75.22)s faster when viewing in 3D than with 2D ( $t(63) = 13.82$ ,  $P < 0.001$ ). At 3 m, the performance advantage of 3D images was greater, and participants were 90.35 (77.74 to 102.96)s faster when using 3D ( $t(63) = 14.32$ ,  $P < 0.001$ ). In relative terms, completion times were 64 per cent shorter at 1 m and 62 per cent shorter at 3 m when viewing in 3D *versus* 2D (calculated as above, using the values presented in Fig. 3b).

#### Suturing

ANOVA on the suturing task performance data yielded two significant effects. Participants completed the task significantly faster when viewing in 3D *versus* 2D

( $F(1,63) = 143.47$ ,  $P < 0.001$ ), and there was also a main effect of viewing distance such that participants were significantly faster at 1 m than at 3 m ( $F(1,63) = 64.87$ ,  $P < 0.001$ ) (Fig. 3c). In relative terms, completion times were 34 per cent shorter at 1 m and 28 per cent shorter at 3 m when viewing in 3D *versus* 2D (calculated as above, using the values presented in Fig. 3c).

### Laparoscopic workload

#### Peg transfer task

ANOVA on the NASA TLX scores for the peg transfer task yielded two significant effects. There was a main effect of viewing mode such that perceived workload was significantly lower when viewing in 3D than in 2D ( $F(1,63) = 90.89$ ,  $P < 0.001$ ). There was also an effect of viewing distance such that perceived workload was significantly lower at 1 m than at 3 m ( $F(1,63) = 26.28$ ,  $P < 0.001$ ) (Fig. 4a). In relative terms, perceived workload was 28 per cent lower at 1 m and 24 per cent lower at 3 m when viewing in 3D *versus* 2D (calculated using the values presented in Fig. 4a, by deducting 3D workload from 2D workload and then dividing by 2D workload for each viewing distance).

#### Navigating in space

Analysis of the NASA TLX scores for the navigating in space task also yielded significant main effects of viewing mode and viewing distance. Perceived workload was significantly lower when viewing in 3D ( $F(1,63) = 131.88$ ,  $P < 0.001$ ) and when viewing from 1 m compared with 3 m ( $F(1,63) = 46.37$ ,  $P < 0.001$ ) (Fig. 4b). In relative terms, perceived workload was 34 per cent lower at 1 m and 32 per cent lower at 3 m when viewing in 3D *versus* 2D (calculated as above, using the values presented in Fig. 4b).

### Suturing

Analysis of the NASA TLX scores for the suturing task again yielded significant main effects of viewing mode and viewing distance. Perceived workload was significantly lower when viewing in 3D ( $F(1,63)=113.80$ ,  $P<0.001$ ), and when viewing from 1 m compared with 3 m ( $F(1,63)=12.03$ ,  $P<0.001$ ) (Fig. 4c). In relative terms, perceived workload was 20 per cent lower at 1 m and 15 per cent lower at 3 m when viewing in 3D versus 2D (calculated as above, using the values presented in Fig. 4c).

### Correlation of laparoscopic performance with Randot® stereoacuity

Across the 12 correlation coefficients assessing the relationship between Randot® stereoacuity and laparoscopic performance on each of the three tasks for each combination of viewing mode (3D, 2D) and viewing distance (1 m, 3 m), there was one significant result after applying the Bonferroni correction (Table 2). Higher Randot® stereoacuity was associated with faster completion times when participants performed the navigating in space task in 3D at 1 m ( $r = -0.40$ ,  $P = 0.001$ ).

### Discussion

There is conflicting evidence about whether the perception of artificial 3D volume translates directly into improved laparoscopic performance<sup>2</sup>. Performance differences identified in previous studies from the operating theatre and simulation laboratory may depend on basic perceptual processes that were not assessed directly, such as individual differences in stereoacuity and tolerance for cross-talk<sup>2</sup>. Both factors were taken into account in the present study by measuring individual stereoacuity and ensuring optimal viewing conditions for the 3D displays in all conditions.

The precision of depth judgements advantage in 3D over 2D is immediate and not affected by either experience or technical differences between novices and experts<sup>1</sup>. However, depth specified by indirect cues in conventional 2D laparoscopy is not always appreciated by inexperienced viewers<sup>1</sup> and must, therefore, be acquired from laparoscopic experience. The Howard–Dolman apparatus used in this study conceptually represents a difficult operating interface that provides minimal indirect depth cues, as encountered in the depths of a human pelvis or within dark, blood-stained tissue. Using this apparatus, the present study showed that laparoscopic depth judgements can be 6.2 times as precise at 1 m, and 3.0 times as precise at 3 m, using 3D compared with 2D displays. These results illustrate large gains in precision that both

novice and experienced surgeons can expect to obtain using 3D displays in challenging locations with minimal indirect depth cues. Given the high risk of intraoperative complications during procedures performed by inexperienced laparoscopists<sup>21–25</sup> and the precision data reported here, current evidence supports the integration of 3D displays into trainees' early operative experiences and training to reduce the risk of these complications arising from imprecise depth judgements. Similarly, in the context of simulated colonoscopy, 3D displays have been shown to improve immediately the detection of diminutive, minimally elevated lesions by trainee endoscopists<sup>26</sup>. The present accuracy findings also demonstrate the relatively high visual fidelity of modern 3D laparoscopic displays.

This study also compared laparoscopic technical performance and perceived workload using 2D and 3D laparoscopic displays, while controlling for viewing distance and optimizing stereopsis by minimizing cross-talk during 3D laparoscopy. Previous clinical and simulation trials<sup>27–30</sup> that did not incorporate these controls found no significant performance differences between 2D and 3D viewing conditions. However, the present study demonstrates that modern 3D displays can offer substantial performance and workload improvements. When participants were tested in the 3D viewing mode, both laparoscopic completion times and perceived workload decreased for all three tasks. As conventional 2D displays do not provide a compelling and immediate perception of 3D volume, this may explain why participants perceived their workload to be higher when viewing in 2D (which required them to extract indirect visual cues while simultaneously performing technical tasks).

Previous studies comparing 3D and 2D displays have used technical tasks proven to correlate with laparoscopic performance in 2D (such as Fundamentals of Laparoscopic Surgery tasks, McGill Inanimate System for Training and Evaluation of Laparoscopic Skills tasks, and the European Training in Basic Laparoscopic Urological Skills system)<sup>22,24,27,28,31–34</sup>. The present study, however, highlights that, for different tasks, laparoscopic performance can vary across different display modes and viewing distances. For the peg transfer task, 3D viewing significantly improved performance, but there was no effect of viewing distance. The relatively small reduction in completion time (13 per cent, averaged over viewing distances) suggests that this task does not require highly precise depth perception. For the suturing task, both 3D viewing and a shorter viewing distance (1 m versus 3 m) significantly improved performance. The larger reduction in completion time associated with 3D viewing (31 per cent, averaged over viewing distances) suggests that this task requires higher



levels of precision than the peg transfer task. Suturing is highly reliant on technique. Both needle and suture are much finer than the plastic triangles used for peg transfer; hence, more precise depth judgements are required to manipulate these objects in 3D space.

In contrast to the peg transfer and suturing tasks, the navigating in space task was designed purposefully to minimize secondary depth cues (by using target loops with minimal contrast and colour variation) and maximize the need for precise depth judgements (by requiring participants to thread a needle through five 2-mm wide target loops arranged differently in 3D space), while minimizing potential technique-related floor effects for laparoscopic beginners (because completing this task is not dependent on esoteric laparoscopic skills). Like the suturing task, both 3D viewing and a shorter viewing distance significantly improved performance. The navigating in space task also demonstrated that the effect of viewing mode on laparoscopic performance can depend on viewing distance, with 3D yielding a significantly greater advantage at 3 m than at 1 m (90.35 versus 65.72 s), although in percentage terms the reductions in completion time compared with 2D were similar (62 per cent at 3 m and 64 per cent at 1 m). These large improvements suggest that precise depth judgements made performance of the task substantially more efficient. Using the navigating in space task at 1 m, a significant correlation between stereoacuity and 3D laparoscopic performance was seen. Before this study, there was no simulation task designed specifically to test performance differences between 2D and 3D displays<sup>2</sup>. Future simulation studies on 3D displays should include manipulations (such as the navigating in space task) that are capable of testing fine levels of laparoscopic depth perception.

The study has limitations, including the high proportion of participants with excellent stereoacuity. This may explain why performance differences in 3D were affected by individual stereoacuity only in the navigating in space task (which required the highest level of precision) and only at the shorter viewing distance (where stereoacuity is maximal). Another explanation as to why only one correlation was found is that participants were beginners, so their limited laparoscopic skills may have made them unable to capitalize on the potential performance advantages of high versus adequate stereoacuity when performing the peg transfer and suturing tasks in 3D.

Strict control over experimental conditions has demonstrated that 3D displays offer significant advantages over 2D displays in terms of precision of depth judgements, laparoscopic performance and perceived workload.

## Acknowledgements

The authors thank Olympus Australia for the loan of laparoscopic displays for independent research. S.S. gratefully acknowledges financial assistance from the Australian Postgraduate Award and the Royal Australasian College of Surgeons Foundation for Surgery PhD Scholarships.

*Disclosure:* The authors declare no conflict of interest.

## References

- Grove PM. The psychophysics of binocular vision. In *3D-TV System with Depth-Image-Based Rendering*, Zhu C, Zhao Y, Yu L, Tanimoto M (eds), vol. 1. Springer: New York, 2012; 347–373.
- Sakata S, Watson MO, Grove PM, Stevenson AR. The conflicting evidence of three-dimensional displays in laparoscopy: a review of systems old and new. *Ann Surg* 2016; **263**: 234–239.
- Sakata S, Grove PM, Stevenson AL. Effect of 3-dimensional vision on surgeons using the Da Vinci robot for laparoscopy: more than meets the eye. *JAMA Surg* 2016; **151**: 793–794.
- Hess RF, To L, Zhou J, Wang G, Cooperstock JR. Stereo vision: the haves and have-nots. *i-Perception* 2015; **6**: 2041669515593028.
- Fergo C, Burcharth J, Pommergaard HC, Rosenberg J. Age is highly associated with stereo blindness among surgeons: a cross-sectional study. *Surg Endosc* 2016; **30**: 4889–4894.
- Sakata S, Grove PM, Hill A, Watson MO, Stevenson AR. The viewpoint-specific failure of modern 3D displays in laparoscopic surgery. *Langenbecks Arch Surg* 2016; **401**: 1007–1018.
- Woods AJ. Crosstalk in stereoscopic displays: a review. *J Electron Imaging* 2012; **21**: 040902.
- Garnham L, Sloper JJ. Effect of age on adult stereoacuity as measured by different types of stereotest. *Br J Ophthalmol* 2006; **90**: 91–95.
- Chandrakumar M, Colpa L, Reginald YA, Goltz HC, Wong AMF. Measuring contrast sensitivity using the M&S Smart System II versus the Pelli–Robson chart. *Ophthalmology* 2013; **120**: 2160–2161.
- Bailey IL, Lovie JE. The design and use of a new near-vision chart. *Am J Optom Physiol Opt* 1980; **57**: 378–387.
- Bailey IL, Lovie-Kitchin JE. Visual acuity testing. From the laboratory to the clinic. *Vision Res* 2013; **90**: 2–9.
- Howard IP. Sensory coding. In *Binocular Vision and Stereopsis*, Howard IP (ed.). Oxford University Press: New York, 1995; 69–104.
- Rubio S, Diaz E, Martin J, Puente JM. Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. *Appl Psychol* 2004; **53**: 61–86.
- Hart SG, Staveland LE. *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*. Amsterdam: Elsevier, 1988.

- 15 Caldiroli D, Molteni F, Sommariva A, Frittoli S, Guanziroli E, Cortellazzi P. Upper limb muscular activity and perceived workload during laryngoscopy: comparison of Glidescope(R) and Macintosh laryngoscopy in manikin: an observational study. *Br J Anaesth* 2014; **112**: 563–569.
- 16 France DJ, Levin S, Hemphill R, Chen K, Rickard D, Makowski R *et al.* Emergency physicians' behaviors and workload in the presence of an electronic whiteboard. *Int J Med Inform* 2005; **74**: 827–837.
- 17 Hubert N, Gilles M, Desbrosses K, Meyer JP, Felblinger J, Hubert J. Ergonomic assessment of the surgeon's physical workload during standard and robotic assisted laparoscopic procedures. *Int J Med Robot* 2013; **9**: 142–147.
- 18 Mohamed R, Raman M, Anderson J, McLaughlin K, Rostom A, Coderre S. Validation of the National Aeronautics and Space Administration Task Load Index as a tool to evaluate the learning curve for endoscopy training. *Can J Gastroenterol Hepatol* 2014; **28**: 155–159.
- 19 Jung SW. A modified model of the just noticeable depth difference and its application to depth sensation enhancement. *IEEE Trans Image Process* 2013; **22**: 3892–3903.
- 20 Jung SW, Ko SJ. Depth sensation enhancement using the just noticeable depth difference. *IEEE Trans Image Process* 2012; **21**: 3624–3637.
- 21 Smith R, Schwab K, Day A, Rockall T, Ballard K, Bailey M *et al.* Effect of passive polarizing three-dimensional displays on surgical performance for experienced laparoscopic surgeons. *Br J Surg* 2014; **101**: 1453–1459.
- 22 Smith R, Day A, Rockall T, Ballard K, Bailey M, Jourdan I. Advanced stereoscopic projection technology significantly improves novice performance of minimally invasive surgical skills. *Surg Endosc* 2012; **26**: 1522–1527.
- 23 Cicione A, Autorino R, Breda A, De Sio M, Damiano R, Fusco F *et al.* Three-dimensional *vs* standard laparoscopy: comparative assessment using a validated program for laparoscopic urologic skills. *Urology* 2013; **82**: 1444–1450.
- 24 Lusch A, Bucur PL, Menhadji AD, Okhunov Z, Liss MA, Perez-Lanzac A *et al.* Evaluation of the impact of three-dimensional vision on laparoscopic performance. *J Endourol* 2014; **28**: 261–266.
- 25 Tanagho YS, Andriole GL, Paradis AG, Madison KM, Sandhu GS, Varela JE *et al.* 2D *versus* 3D visualization: impact on laparoscopic proficiency using the fundamentals of laparoscopic surgery skill set. *J Laparoendosc Adv Surg Tech A* 2012; **22**: 865–870.
- 26 Sakata S, Grove PM, Stevenson ARL, Hewett DG. The impact of three-dimensional imaging on polyp detection during colonoscopy: a proof of concept study. *Gut* 2016; **65**: 730–731.
- 27 Alaraimi B, El Bakbak W, Sarker S, Makkiyah S, Al-Marzouq A, Goriparthi R *et al.* A randomized prospective study comparing acquisition of laparoscopic skills in three-dimensional (3D) *vs.* two-dimensional (2D) laparoscopy. *World J Surg* 2014; **38**: 2746–2752.
- 28 Mistry M, Roach VA, Wilson TD. Application of stereoscopic visualization on surgical skill acquisition in novices. *J Surg Educ* 2013; **70**: 563–570.
- 29 Kyriazis I, Ozsoy M, Kallidonis P, Vasilas M, Panagopoulos V, Liatsikos EN. Integrating three-dimensional vision in laparoscopy: the learning curve of an expert. *J Endourol* 2015; **29**: 657–660.
- 30 Kinoshita H, Nakagawa K, Usui Y, Iwamura M, Ito A, Miyajima A *et al.* High-definition resolution three-dimensional imaging systems in laparoscopic radical prostatectomy: randomized comparative study with high-definition resolution two-dimensional systems. *Surg Endosc* 2015; **29**: 2203–2209.
- 31 Patel HR, Ribal MJ, Arya M, Nauth-Misir R, Joseph JV. Is it worth revisiting laparoscopic three-dimensional visualization? A validated assessment. *Urology* 2007; **70**: 47–49.
- 32 Bhayani SB, Andriole GL. Three-dimensional (3D) vision: does it improve laparoscopic skills? An assessment of a 3D head-mounted visualization system. *Rev Urol* 2005; **7**: 211–214.
- 33 Herron DM, Lantis JC II, Maykel J, Basu C, Schwaitzberg SD. The 3-D monitor and head-mounted display. A quantitative evaluation of advanced laparoscopic viewing technologies. *Surg Endosc* 1999; **13**: 751–755.
- 34 Wagner OJ, Hagen M, Kurmann A, Horgan S, Candinas D, Vorburger SA. Three-dimensional vision enhances task performance independently of the surgical method. *Surg Endosc* 2012; **26**: 2961–2968.