Memory and Metacognition in Dangerous Situations: Investigating Cognitive Impairment From Gas Narcosis in Undersea Divers

Malcolm Hobbs, University of Winchester, Winchester, UK, **Philip A. Higham**, University of Southampton, Southampton, UK, and **Wendy Kneller**, University of Winchester, Winchester, UK

Objective: The current study tested whether undersea divers are able to accurately judge their level of memory impairment from inert gas narcosis.

Background: Inert gas narcosis causes a number of cognitive impairments, including a decrement in memory ability. Undersea divers may be unable to accurately judge their level of impairment, affecting safety and work performance.

Method: In two underwater field experiments, performance decrements on tests of memory at 33 to 42 m were compared with self-ratings of impairment and resolution. The effect of depth (shallow [1-11 m] vs. deep [33-42 m]) was measured on free-recall (Experiment 1; n = 41) and cued-recall (Experiment 2; n = 39) performance, a visual-analogue self-assessment rating of narcotic impairment, and the accuracy of judgements-of-learning (JOLs).

Results: Both free- and cued-recall were significantly reduced in deep, compared to shallow, conditions. This decrement was accompanied by an increase in self-assessed impairment. In contrast, resolution (based on JOLs) remained unaffected by depth. The dissociation of memory accuracy and resolution, coupled with a shift in a self-assessment of impairment, indicated that divers were able to accurately judge their decrease in memory performance at depth.

Conclusion: These findings suggest that impaired self-assessment and resolution may not actually be a symptom of narcosis in the depth range of 33 to 42 m underwater and that the divers in this study were better equipped to manage narcosis than prior literature suggested. The results are discussed in relation to implications for diver safety and work performance.

Keywords: memory, inert gas narcosis, metacognition, diving, judgements of learning, free recall, cued recall, resolution, diver safety

Address correspondence to Wendy Kneller, Department of Psychology, University of Winchester, Winchester, SO22 4NR, UK; e-mail: wendy.kneller@winchester.ac.uk.

HUMAN FACTORS
Vol. XX, No. X, Month 2013, pp. 1–14
DOI: 10.1177/0018720813510737

Understanding exactly when and how human performance is impaired in potentially dangerous situations can lead to increases in safety, work performance, and ultimately the development of methods to reduce or eliminate the impairment. For example, identifying which cognitive skills are impaired in such diverse situations as drug and alcohol intoxication (Kelly, Darke, & Ross, 2004), high-altitude mountaineering (Lieberman, Protopapas, & Kanki, 1995), and phone use while driving (Alm & Nilsson, 1994) have led to changes in training practices, legislation, and safety guidelines. However, as important as identifying specific impairments may be in these situations, it is arguably just as important to understand whether individuals are aware of and able to accurately judge impairment in themselves. That is, what are an individual's *metacognitive skills* (i.e., skills related to the assessment of one's own cognition) in relation to impaired performance on a task?

Although theorists have addressed a number of important research questions about metacognition (defined as cognition about cognition) in a variety of domains including education and forensics, among others, little research has focused on metacognition in potentially dangerous situations and environments. However, it is important to investigate this relationship because the little research that has been conducted has suggested that it is possible for monitoring of performance (metacognitive awareness) and cognitive performance (specific cognitive impairment) to dissociate under certain conditions. For example, Nelson et al. (1990) reported that hypoxia in mountain climbers on Everest degraded metacognition independently of recognition memory ability. Similarly, it has been reported that alcohol can affect measures of memory ability but not metacognition (Curran & Hildebrandt, 1999; Nelson, McSpadden, Fromme, & Marlatt, 1986). Understanding the relationship between memory ability and metacognition is particularly important in these situations because if individuals are unable to accurately assess whether they are impaired or not, it can impact upon decisions to engage in compensatory strategies and behaviors to avert accidents and poor work performance. This may lead people to attempt tasks on which they do not realize they are impaired or forget critical information related to safety that could lead to a serious or fatal accident. Another situation where similar effects may prevail is in the case of inert gas narcosis, experienced in pressurized environments.

Inert Gas Narcosis

Exposure to increased ambient pressure leads to the phenomenon known as inert gas narcosis (see Bennett & Rostain, 2002, for review). Narcosis is defined as a spectrum of cognitive impairments and behavioral alterations, including euphoria, light-headedness, clumsiness, impaired memory, and at extreme pressures, disorientation and loss of consciousness (De Gorordo, Vallejo-Manzur, Chanin, & Varon, 2003). Narcosis is caused by the presence of inert gases in breathing mixtures and is often referred to as nitrogen narcosis after the gas held primarily responsible when breathing normal air (21% oxygen, 79% nitrogen). Encountered in any pressurized environment, narcosis is most often experienced by undersea divers, where symptoms become apparent at pressures of four atmospheres (ata; equivalent to 30 m underwater) when breathing normal air. Narcosis becomes progressively worse with increases in pressure but can be delayed until higher pressures using alternative breathing mixtures, such as those replacing nitrogen with helium (Logie & Baddeley, 1985). The concerns for individuals in high-pressure environments are that narcosis is known to impair work performance (Morisson & Zander, 2008; Pulak, 1998; Van Rees Vellinga, Verhoeven, Van Dijik, & Sterk, 2006) and to be a factor in deep diving related accidents (Hart, White, Conboy, Bodiwala, & Quinton, 1999) and fatalities (Edmonds & Douglas, 1990; Levett & Millar, 2008).

The most evident symptoms of narcosis have been well established in studies showing that cognitive and motor performance on a wide variety of tasks is measurably impaired from 30 m/4 ata (Baddeley, De Figueredo, Hawkswell-Curtis, & Williams, 1968; Kiessling & Maag, 1962; Tetzlaff et al., 1998). For example, free-recall performance is reduced at depths of 35 m/4.5 ata (Fowler, 1973; Hobbs & Kneller, 2009; Kneller & Hobbs, 2013; Philp, Fields, & Roberts, 1989; Tetzlaff et al., 1998), and Fowler, White, Wright, and Ackles (1980) have suggested that cued-recall performance is affected at these depths as well.

The literature (e.g., Gronning & Aarli, 2011) and training organizations (e.g., Shreeves, 2008; Strauss & Akensov, 2004) also list more intangible symptoms of narcosis, such as loss of judgment, overconfidence, and irrational behavior. The concern is that cognitive/motor impairment coupled with a concurrent loss of judgment generates a situation where individuals affected by narcosis are unaware of, or are unable to accurately assess, their own level of impairment. This may result in divers overestimating their abilities on an underwater task, failing to react appropriately in dangerous situations, and setting in motion a chain of events that leads to a serious diving accident.

Self-Assessment

Clearly, this is a situation to be avoided and it highlights the fact that accurate self-assessment of cognitive impairment is almost as critical as the impairment itself. However, reliable evidence pertaining to divers' self-assessment is lacking and the little available is contradictory. For example, using just 11 divers, Hamilton, Laliberte, and Fowler (1995) reported a dissociation between a subjective estimate of the strength of narcosis and performance on a reaction time (RT) task, suggesting poor self-assessment of ability. However, subjective ratings of work effectiveness recorded concurrently paralleled performance on the RT task, suggesting that the divers possessed the ability to judge their own performance accurately. The authors have also found previously that when running field studies (e.g., Hobbs, 2008; Hobbs & Kneller, 2009), divers often anecdotally report on the surface that they are not personally affected by narcosis, have become tolerant to its effects, or claim never to have experienced narcosis even at depths of 60 m/7 ata on normal air. On the other hand, when engaged in an actual dive, the story was different: divers indicated an increase in subjective impairment using simple rating scales with a concurrent drop in task performance.

Metacognitive Resolution

An issue related to self-assessment of impairment is metacognitive *resolution*, discussed predominantly in the literature on human memory (e.g., Brewer, Keast, & Rishworth, 2002; Higham & Arnold, 2007). Resolution refers to people's ability to assess the correctness of their own responses. Unlike self-assessment, which is a global judgment about the current state of cognitive functioning, resolution is typically measured across multiple trials.

To index resolution, either prospective (i.e., how good performance will be in the future) or retrospective (i.e., how good performance was in the past) metacognitive ratings are needed. For example, participants may be asked to study cue-target pairs and then judge prospectively (at study) how likely (0%−100%) it is that they would be able to remember a target word when prompted with a cue word in a test 10 minutes from now, a rating referred to as a *judgment-of-learning* (JOL; e.g., Hanczakowski, Zawadzka, Pasek, & Higham, in press; Koriat, 1997). Conversely, participants may be asked to judge retrospectively how likely (0%−100%) it is that an item just recalled on the test is actually a target, a rating referred to as *retrospective confidence*. Once the metacognitive ratings have been obtained, there are various methods that can be used to determine how well those ratings discriminate response correctness ranging from simple computation of a correlation coefficient to more sophisticated methods involving signal detection theory. The latter method, detailed in the following, is the one we adopted in the present work.

Experimental Overview

The primary aims of the current two studies were to investigate whether divers' recall performance and/or resolution was impaired at deep depths (33–42 m/4.3–5.2 ata), compared to shallow depths (<12 m/2.2 ata). These depths were chosen because the general consensus (e.g., Bennett & Rostain, 2002) is that narcosis becomes apparent and measurable from depths of 30 m/4 ata, whereas it does not appear to manifest shallower than 12 m/2.2 ata. A shallow water control, rather than a surface-based "dry" control, was initially decided upon to control for any impairment caused by merely being in the water (e.g., Ross, 1989) or for any environmental context effects (e.g., Godden & Baddeley, 1975) known to affect memory performance. Thus, any effect of depth could be reasonably attributed to narcosis, rather than differences in the testing environment. In Experiment 1, a free-recall test was used, whereas the test in Experiment 2 was cued-recall for reasons outlined in the following. To investigate resolution in each experiment, both JOLs and retrospective confidence ratings were collected. To date, we know of no published studies investigating the effects of narcosis on resolution. A secondary aim was to investigate whether divers were able to make an accurate, global self-assessment of any impairment that they were suffering. To measure this, we had divers rate their level of impairment on a subjective visual-analogue (V-A) scale.

EXPERIMENT 1: FREE RECALL

As noted previously, narcosis has been shown to reliably impair free recall in previous research (e.g., Hobbs & Kneller, 2009). Hence, a free-recall task was a good place to begin our investigation into the question of whether narcosis has a similar or a dissociative effect on resolution as memory impairment was a likely outcome.

Method

Participants. All participants were customers or staff at the recreational dive facility Nautilus Watersports in Port Vila, Vanuatu, and were screened to ensure that they were suitably qualified and medically fit to dive to the depths required in the study. The study was granted ethical approval by the University of Winchester (UK). Forty-one native English-speaking divers (29 male) volunteered for the study. The preponderance of males in the sample (and in Experiment 2) reflects the larger number of male, compared to female, divers in the recreational diving community (e.g., Hagberg & Ornhagen, 2003). Exploratory analyses revealed no effects of gender on any measure reported in this paper, so questions of gender will therefore not be discussed further. Participants were aged between 18 and 60 years (M = 35.2, SD = 11.9). (There was a large variation in the age of the participants and it is well known that aging can have a detrimental effect on memory [e.g., Craik, 1994]. However, we could find no significant effects of age, and an analysis of the data without participants >50 years of age resulted in no changes to our central conclusions. We therefore chose not to discuss this factor in the interests of clarity and brevity.) Participants held qualifications ranging from PADI advanced open water (or equivalent) up to instructor and technical diving certifications. All participants had completed between 14 and 1,500 (M = 254.6, SD = 347.8) dives over a period of between 0.3 and 26 years (M = 6.96, SD = 6.2). For the purposes of the analyses, participants were assigned experience level by conducting a median split for the number of dives they had completed. Those in the low-experience group had completed between 14 and 83 dives (M = 42.6, SD = 23.9) and those in the high-experience group completed between 85 and 1,500 dives (M = 456.4, SD = 392.0). Once testing was completed on each dive, participants were given a guided tour of the dive site by one of the researchers as their incentive for taking part in the study.

Design and measures. The experiment was designed so that we could compare (within participants) the effect of depth (shallow vs. deep) on free-recall performance, subjective ratings of impairment (V-A scale), and resolution. In the shallow condition, participants were tested at depths of 1 m to 11 m (M = 1.24, SD = 1.6) and in the deep condition at depths of 33 m to 42 m (M = 37.1, SD = 2.2). In addition to depth, the effects of three between-participants variables that may also have influenced the results were also considered. These were: depth order, list order, and experience level (low experience vs. high experience). The depth-order and list-order variables were created through counterbalancing as explained in the following.

The free-recall memory test consisted of two wordlists (A and B), each containing 15 words. All words were matched for concreteness and imageability using the MRC linguistic database (v 2.0). Each participant was presented with one list at each depth (shallow [S] and deep [D]). For the first 40 participants, the order of presentation of the lists (A then B or B then A) was counterbalanced across the depth-order conditions (shallow-deep or deep-shallow) so that there were an equal number of participants (n = 10) in each of the four possible combinations: shallow-deep-AB, shallow-deep-BA, deep-shallow-AB, deep-shallow-BA. The 41st participant was assigned to the deep-shallow-BA combination.

Participants also rated each word, as it was presented for study, for the likelihood (0%–100%) that they would be able to remember it in the recall phase (JOL). Participants also provided a retrospective confidence rating (0%–100%) in the test phase as to how confident they were that each word they recalled was correct. The visual-analogue scale of impairment consisted of a rating of believed impairment from narcosis along a 100 mm line with the anchors *not impaired* and *extremely impaired* and a central mark at 50 mm. The scale was printed on an underwater slate and participants responded by making a mark along the scale. The point at which their mark intersected the line was measured in millimeters and constituted their score.

Procedure. Participants were briefed that the aim of the study was to test their memory ability on deep dives. Participants completed the shallow and deep conditions as two separate sessions on the same day with between 1 and 3 hours between each session. The shallow conditions were conducted either in a swimming pool (n = 40), in full dive equipment, or in the ocean (n = 1), and the deep conditions were all conducted in the ocean. Using both ocean and pool environments for the shallow conditions was deemed acceptable as prior evidence (e.g., Kneller & Hobbs, 2013) has shown that tropical ocean and swimming pool environments do not lead to a difference in memory ability. Depth was measured by holding a dive computer at chest height. Ocean sessions were conducted either as a shore dive or from a boat. Six different but topographically similar ocean sites in Mele Bay, Vanuatu, were used for data collection. Each site consisted of a sloping reef table with a flat or gently sloping sandy or dead coral seabed between 30 m and 45 m. On the shore dives, the required depth could be reached within a short distance and with the minimum swimming out to sea. This caused a level of exertion no different from diving from the boat. Ocean conditions were stable with little or no current/surge on the seabed, a consistent temperature of 28°C to 29°C in the ocean, and visibility of 10 m to 30 m (dependent on site and daily conditions). The temperature of the pool was 27°C. All dives were conducted on normal air (21% oxygen, 79% nitrogen).

Participants were briefed in detail on the protocol on the surface before going underwater to ensure they fully understood the tasks. They were shown dummy example words printed on laminated cards the same as the real test stimuli and the underwater writing slates they would be using to record all their responses. The slates were clearly labeled with columns and sections for them to fill in the relevant JOLs, recalled words, and so on. Instructions were also written on each slate to remind them of what to do. Additional instructions underwater were given using recognized diving-related hand signals (e.g., “OK”, “Stop”), and the researcher carried an additional wrist slate for writing messages for the participants to read if needed.

Descent time from surface to required depth in the deep condition was between 2 and 9 minutes. The protocol was identical in both depth conditions. Testing took place immediately upon reaching the required depth and participants were settled on the ocean or swimming pool bottom. Participants were first presented the 15 words on laminated cards by the experimenter. Participants were shown each word for 5 seconds, timed using a stopwatch. While each word was presented participants wrote on a slate how confident they were that they would be able to recall the word in the recall phase (JOL). Once all words had been presented there was a break of 2.5 minutes during which subjects carried out a manual dexterity distractor task and completed the V-A scale of impairment from narcosis. The manual dexterity formed part of another project and is reported elsewhere (Kneller, Higham, & Hobbs, 2012). (The dexterity task consisted of a plastic block with holes for two rows of axles held in place with grommets at either end. Participants were required to transfer the axles and grommets to the other end of the block and back again as quickly as possible throughout the break between presentation and recall.) After the break, while still underwater at the same depth, subjects were handed a blank slate on which they wrote down as many words as they could remember. Immediately after they wrote down each word, they provided a rating next to it as to how confident they were that the word they had written down was correct (retrospective confidence).

Data analyses. The means for the percentage of words recalled and V-A ratings were calculated for the deep and shallow conditions. We analyzed each of the measures with a 2 × 2 × 2 × 2 mixed analysis of variance (ANOVA) with depth (shallow vs. deep) as a within-participants factor and depth order (SD vs. DS), presentation of list order (AB vs. BA), and experience level (low experience vs. high experience) as between-participants factors.

In the metacognitive literature, resolution has typically been measured with the Goodman-Kruskal gamma coefficient (Goodman & Kruskal, 1954), an ordinal measure of association recommended by Nelson (1984). However, recently gamma has been shown to have a number of undesirable properties, and others have suggested using signal detection methodology to index resolution instead (e.g., Higham, 2002, 2007, 2011; Luna, Higham, & Martin-Luengo, 2011; Masson & Rotello, 2009; Rotello, Masson, & Verde, 2008). For our data, we measured resolution by computing the area under the receiver operating characteristic (ROC) curves (AUC; e.g., Green & Moses, 1966; Pollack, Norman, & Galanter, 1964). ROC curves are a plot of the hit rate as a function of the false alarm rate for different levels of the response criterion. In our context, the hit rate and the false alarm rate are, respectively, defined as the

TABLE 1: Means (*SD*s) for Cued Recall, V-A Ratings, AUC, Retrospective Confidence, and JOLs in Experiment 1

	Diving Depth	
Measure Type	Shallow (1–11 m)	Deep (33–42 m)
Main measures		
Mean recall score[a]	45% (12.9)	36% (13.5)
V-A rating[a]	2.22 (6.8)	24.45 (18.7)
Resolution (AUC)	0.60 (0.14)	0.61 (0.16)
Other measures		(9.6)
Retrospective confidence	98.9 (2.4)	95.7 (21.0)
JOL rating	56.9 (17.4)	53.6

Note. V-A = visual-analogue; AUC = area under the curve; JOL = judgment of learning.
[a]Shallow and deep scores were significantly different from each other on this measure.

proportion of correct and incorrect responses assigned a given confidence level or higher. To generate the ROC, various hit and false alarm rates are computed by treating different levels of confidence (10, 20, 30, and so on) as different response criteria. The diagonal of the ROC curve represents an inability to distinguish between correct and incorrect responses at any level of confidence (chance-level resolution: AUC = 0.5). As the curve bows from the diagonal toward the top left of the figure, discrimination (resolution) increases (perfect resolution: AUC= 1.0).

Because JOLs were provided to all targets at study, omitting a response at test could be considered an error with an associated rating. However, retrospective confidence ratings were not provided for response omissions at test and there were very few intrusions (reported errors). Consequently, there were not enough ratings assigned to recall errors to compute resolution on the retrospective measure.

Results and Discussion

Means for the percentage of study items freely recalled, V-A ratings of impairment, and resolution (AUCs based on JOLs) are displayed in Table 1. Means for JOLs and retrospective confidence are also shown in Table 1 for completeness, but as the primary purpose of these measures was to compute resolution, and because retrospective confidence means were near ceiling, differences between means were not statistically analyzed.

There was a decrease in mean recall of 9% in the deep water compared to the shallow water, confirmed by a significant main effect of depth from the four-way ANOVA, $F(1, 33) = 20.21$, $p < .001$. There were no significant main effects of presentation of list order or experience level, both *p*s > .05. There was a significant main effect of depth order, $F(1, 33) = 5.90$, $p = .021$, which indicated that recall generally was lower for participants in the SD ($M = 5.4$, $SD = 2.4$) compared to the DS ($M = 6.7$, $SD = 1.5$) condition. Crucially, there was no significant interaction involving depth, largest $F(1, 33) = 2.83$, $p = .102$, indicating that the effect of depth held regardless of the level of the other factors in the design.

The V-A ratings ranged from 0 to 41 in the shallow water and from 0 to 72 in the deep water. There was a 22.2 point mean increase in perceived impairment in the deep water compared to shallow water. This increase was confirmed by the ANOVA, $F(1, 32) = 67.72$, $p < .001$, indicating that participants believed they were more impaired in the deep water compared to shallow water. There was no other significant main effect or interaction from the four-way ANOVA, all *p*s > .05.

Resolution for JOLs is displayed in Figure 1. Two participants were dropped from the resolution analysis in the deep condition; one failed to recall any items (rendering undefined hit rates)

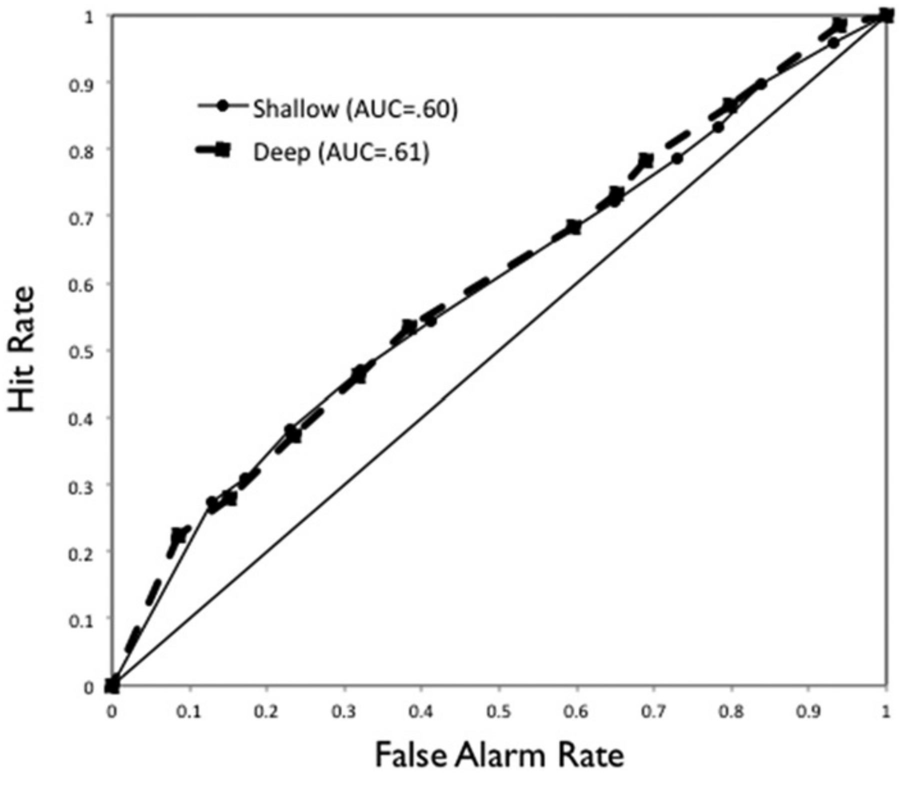

Figure 1. Prospective (JOL-based) resolution for free recall (Experiment 1). The diagonal represents chance-level resolution (inability to distinguish between correct and incorrect responses). The greater the bow of the curve to the top-left of the figure, the greater the ability to distinguish between correct versus incorrect responses (i.e., higher resolution as indexed by higher AUC). JOL = judgment-of-learning; AUC = area under the curve.

and another failed to provide any confidence ratings. Resolution, as measured by AUC, was above chance for both the shallow, $t(38) = 4.29$, $p < .001$, and deep, $t(38) = 4.10$, $p < .001$, conditions. The four-way ANOVA that tested for an effect of depth revealed no main effects or interactions, all *p*s > .05. Indeed, as Figure 1 shows, AUC was remarkably similar across the shallow and deep conditions.

Experiment 1 demonstrated that under narcosis, free-recall performance was impaired and participants accurately self-assessed their impairment as measured by the V-A ratings. However, resolution remained intact across the different depths. Together, the results suggest that although the divers suffered narcosis-based memory impairment and were consciously aware of that impairment (as indicated by V-A ratings), their ability to discriminate between items that they would later recall and those that they would not was unaffected. This is encouraging news for divers; it suggests that memory impairments they may encounter at deep depths will not be accompanied by a comparable decrement in their metacognitive processing. As a result, they may be in a position to take steps to compensate for their reduction in memory ability to minimize any damage that might otherwise occur (e.g., processing critical information for longer periods if it is deemed likely to be forgotten).

EXPERIMENT 2: CUED RECALL

In Experiment 2, we aimed to replicate Experiment 1 and expand upon those results by employing a cued- instead of a free-recall test. Although numerous studies have shown free recall to be affected by narcosis, some other memory measures such as recognition memory have been unaffected (e.g., see Hobbs & Kneller, 2009). A possible reason for the discrepancy is that free recall involves self-initiated retrieval, making it more demanding than recognition on cognitive resources that may be in short supply at deep depths (e.g., Craik & McDowd, 1987). If so, it may be that providing cues at test may reduce the need for self-initiated retrieval, freeing up precious resources and reducing narcosis-based memory impairment. We are aware of only one prior narcosis study that has used cued recall (Fowler et al., 1980), and the authors reported a deleterious effect. Nonetheless, the data bearing on the issue are limited, and given the potential to further our theoretical understanding of the relationship between narcosis and retrieval processes, a cued-recall task was used in Experiment 2.

Method

Participants. Thirty-nine participants (25 male) volunteered for the study, aged between 18 and 60 years (M = 34.41, SD = 12.0), of which 29 had already completed Experiment 1. Some participants took part in both experiments because the data for both had to be collected concurrently in order to meet the logistical and time constraints for collecting the data imposed on the authors. Participants who had completed Experiment 1 took part in Experiment 2 on a different day with at least 24 hours between the end of Experiment 1 and the start of Experiment 2. Participants had completed between 14 and 1,000 (M = 161.2, SD = 222.5) dives over a period of between 0.5 and 30 years (M = 6.38, SD = 6.6). Divers were split into two experience level groups using the median as in Experiment 1. Those in the low experience group had completed between 14 and 75 dives (M = 40.15, SD = 20.3) and those in the high experience group between 83 and 1,000 dives (M = 288.6, SD = 265.8).

Design and measures. The design was exactly the same as Experiment 1, except the free-recall test was replaced with a cued-recall test. The shallow and deep conditions in this case consisted of depths between 1 m and 11 m (M = 1.74, SD = 2.3) and 34 m to 41 m (M = 38.03, SD = 1.9), respectively.

Two wordlists (A and B), each containing 15 word pairs, were generated. These word lists did not contain any of the words from Experiment 1. All words were matched for concreteness and imageability using the MRC linguistic database (v 2.0). The words were randomly paired and those with spurious relationships reassigned to words that had no obvious association in meaning or language usage. The order and number of times each list appeared in each depth condition was counterbalanced. Thus, each depth condition (shallow-deep and deep-shallow) had the AB and BA list order appear 10 times, except for the shallow-deep/AB combination, which only had 9 participants.

Procedure. Exactly the same dive sites were used as in Experiment 1. Four participants completed the shallow conditions in the ocean and 35 in the swimming pool. Descent time to deep depth ranged from 2 to 4 minutes. The procedure was the same as in Experiment 1 except that participants were provided with cue-target word pairs at study and the studied cues at test. During the study phase, participants were informed that they would be shown word pairs and later (in the recall phase) provided with only the first word of each pair to use as a cue to recall the second word. To accommodate the cues for the JOL rating, participants were instructed at study to rate how likely (0%−100%) they would be to recall the second word if they were presented with the first word in the pair during testing. Both the cue and the target were in full view for this JOL rating at study. In the break between presentation and recall, the same manual dexterity task was completed as in Experiment 1.

In the recall phase, participants were handed a slate containing all the cue words. They were instructed to recall and write down next to each cue the target that was paired with it during study. The order of presentation of the cues was different to that of the study phase. For the retrospective confidence ratings made at test, participants rated how confident they were that each word they recalled was correct (0%−100%).

Data analyses. Strict scoring was used, meaning that only words that were correctly reported and paired with their studied cues were counted as correct responses. (The data were also analyzed using more liberal scoring for which targets that were correctly recalled were counted as correct even if they were paired with

TABLE 2: Means (*SD*s) for Cued Recall, V-A Ratings, AUC, Retrospective Confidence, and JOLs in Experiment 2

	Diving Depth	
Measure Type	Shallow (1–11 m)	Deep (34–41 m)
Main measures		
Mean recall score[a]	29% (19.5)	19% (17.5)
V-A rating[a]	1.64 (3.1)	21.9 (17.4)
Resolution (AUC)	0.66 (0.2)	0.74 (0.2)
Other measures		(21.0)
Retrospective confidence	90.0 (13.8)	89.1 (18.5)
JOL rating	45.8 (20.9)	41.5

Note. V-A = visual-analogue; AUC = area under the curve; JOL = judgment of learning.
[a]Shallow and deep scores were significantly different from each other on this measure.

cues that they were not paired with at study. However, scoring this way did not alter the conclusions of the study. The results are thus not reported in the interest of brevity.) One participant's cued-recall data from the deep condition was lost due to a technical difficulty on one of the dives, leaving data from 38 participants available for analysis. As in Experiment 1, the degrees of freedom vary across the statistical tests because of missing data. The data were treated in the same manner as in Experiment 1. That is, cued-recall performance, V-A ratings, and resolution were analyzed with a 2 × 2 × 2 × 2 mixed ANOVA with depth, presentation of word list order, depth order, and experience level as factors. Only depth was varied within participants.

Results and Discussion

Table 2 displays the means for the percentage of targets produced in cued recall, V-A ratings of impairment, and resolution (AUCs based on JOLs). As in Experiment 1, means for JOLs and retrospective confidence are also included in Table 2 for completeness, but they were not statistically analyzed.

There was a significant, 10% mean decrease in cued recall in the deep condition compared to the shallow condition, $F(1, 30) = 5.87$, $p = .022$. There were no other significant main effects or interactions from the four-way ANOVA, all ps > .05.

The V-A ratings ranged from 0 to 13 in the shallow water and 0 to 65 in the deep water. The four-way ANOVA on these ratings indicated that the impairment to cued recall from depth just reported was accompanied by a significant 20.3 point mean increase in the V-A ratings, $F(1, 31) = 53.09$, $p < .001$, demonstrating that participants believed they were more impaired in the deep water than in the shallow water. There was also a main effect of list order, $F(1, 31) = 4.37$, $p = .045$, indicating that ratings for the AB presentation order ($M = 13.89$, $SD = 17.8$) were higher than the BA presentation order ($M = 9.75$, $SD = 14.1$). There was no other significant main effect or interaction from the four-way ANOVA, all ps > .05.

Resolution is displayed in Figure 2. Eight participants were dropped from the deep condition and one from the shallow condition for failing to recall any targets, which rendered undefined hit rates. (It was a concern that excluding so many participants, who also happened to be the poorest recall performers, altered the sample such that it became very different from that used for the recall data. We therefore conducted another analysis of the recall data in which participants whose resolution data was excluded also had their recall data excluded. That way, exactly the same participants were analyzed for each measure and all recalled at least some targets. As the initial analysis showed no effect of list order, depth order, or experience, we simplified the new analysis to a *t*-test, which revealed

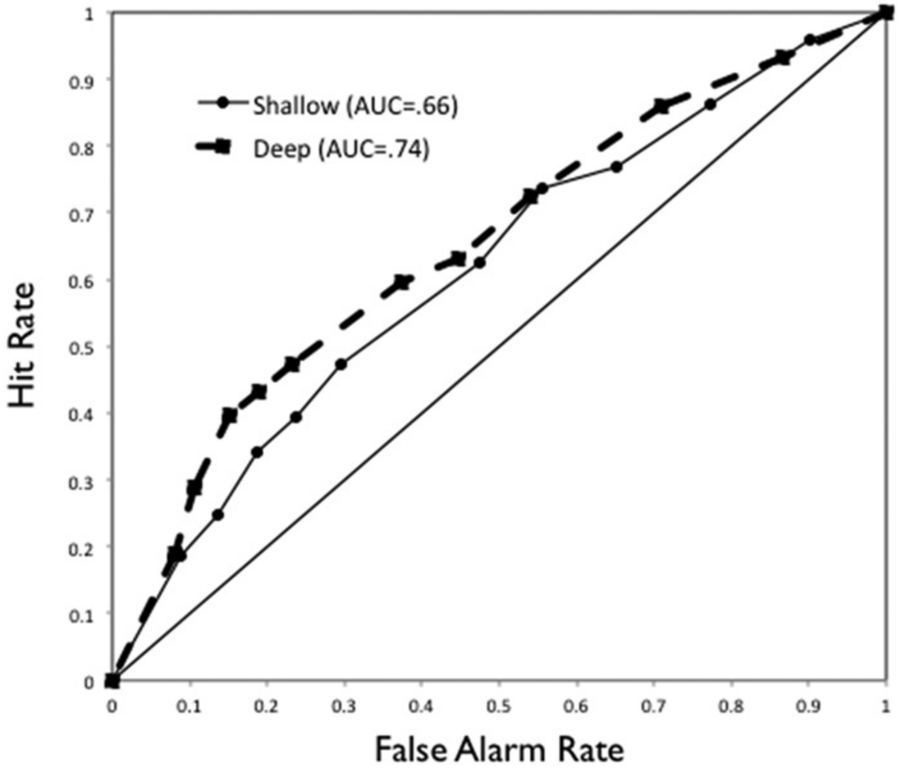

Figure 2. Prospective (JOL-based) resolution for cued recall (Experiment 2). JOL = judgment-of-learning; AUC = area under the curve.

a significant decrease in cued-recall performance due to depth, $t[28] = 2.21$, $p = .036$. This result replicates the depth-induced dissociative pattern between recall and resolution reported in the main analyses.) Resolution (AUC) was above chance at both shallow, $t(37) = 4.89$, $p < .001$, and deep, $t(28) = 6.88$, $p < .001$, depths. Interestingly, mean resolution was numerically *higher* in the deep condition ($M = .74$, $SD = 0.2$) than in the shallow condition ($M = .66$, $SD = 0.2$), although the four-way ANOVA indicated that this difference was not significant, $F < 1$. No main effects of depth order, list order, or experience level were found, all *p*s > .05. However, significant interactions were found between experience and list order, $F(1, 21) = 6.31$, $p = .02$, and between depth order and list order, $F(1, 21) = 5.53$, $p = .029$. As these interactions do not involve depth, the variable of main interest, they are not of critical importance. No other interaction from the four-way ANOVA was significant.

The results of Experiment 2 replicated those from Experiment 1. Memory accuracy (cued-recall performance) was impaired at depth, and participants reported that they were aware of this impairment (V-A ratings). In contrast, resolution remained intact at depth. The resolution data indicate that despite their cognitive deficit, participants were still able to discriminate items that they would later be able to recall from those that they could not at the same level in shallow versus deep water.

Cued-recall performance in Experiment 2 was poor in both the shallow and deep conditions, even poorer than in Experiment 1 where no cues were provided. The reason for this poor performance is not immediately clear and differences between free and cued recall in underwater environments might be addressed in future research. Perhaps the unusual context of the study meant that the requirement to use the experimenter-provided cues was particularly difficult. Nonetheless, the results clearly suggest that the effect of narcosis on free recall (in Experiment 1 and in previous research) was not due to self-initiated retrieval processes in free recall drawing on cognitive resources that were limited by an underwater context. Narcosis deteriorated recall performance in Experiment 2 even though cues were provided at test to reduce the need for self-initiated retrieval.

One limitation of Experiment 2 is that most of the participants (by practical necessity) had recently completed Experiment 1, which may have constituted a form of practice that may have influenced the metacognitive judgments in Experiment 2. In contrast, participants in Experiment 1 were completing the task underwater for the first time. Future research might compare free- and cued-recall performance between different participant samples to alleviate this concern.

GENERAL DISCUSSION

In two experiments, performance on both free- and cued-recall tests was significantly worse at deep depths (33–42 m/4.3–5.2 ata), compared to shallow depths (1–11 m/1.1–2.1 ata). This finding indicated an impairment of memory performance at depths considered to be narcotic and replicated previous studies that have found the same decrement in memory performance (e.g., Hobbs & Kneller, 2009; Philp et al., 1989). In both experiments, participants reported that they believed to be experiencing some degree of impairment from narcosis as measured by the V-A scale. In contrast, resolution in both experiments did not differ between deep and shallow depths, showing no sign at all of impairment. In fact, there was a numeric (but not statistical) resolution *advantage* in the deep compared to the shallow condition in Experiment 2. Together, these findings add to the literature from other situations demonstrating that metacognitive awareness and cognitive performance can become dissociated, such as under the effects of alcohol (Curran & Hildebrandt, 1999) and hypoxia (Nelson et al., 1990). These findings also highlight that there are other potentially dangerous situations where it would be useful in future research to assess the relationship between performance impairment and metacognitive awareness. Such situations might include mental performance in astronauts (Manzey, 2000; Manzey, Lorenz, Heuer, & Sangals, 2000) or fatigue while operating vehicles (e.g., trains: Dorrian, Roach, Fletcher, & Dawson, 2007).

Remembering lists of words is an unusual task for divers, and future studies may seek to use tasks that are more relevant to diving. Even so, the current study has implications for underwater safety and work performance. First, the deficits we observed in recall could generalize to a failure to remember important information at the appropriate moment during a dive. Such memory failures may result in a delay in responding to a problem that could lead to injury or a fatality. However, being able to accurately judge which information will be later remembered and which will not potentially allows divers to compensate for any narcotic memory impairment that they experience. For example, divers who are engaged in a deep dive may have to judge whether they will later remember to check or switch air supply, complete decompression stops, or think twice about entering caves or wrecks where the exit might be hard to locate. If they correctly judge that this information will be forgotten when at a deep depth (good resolution), they might create a reminder for themselves, process the information for a longer period, or ask fellow divers to remind them of the information later when it is needed.

However, it is also important to emphasize that monitoring and control are different constructs (e.g., Nelson & Narens, 1990, 1994). Our research has shown that one type of monitoring accuracy—resolution—is intact at deep depths. However, is divers' ability to implement compensatory control procedures in response to that accurate monitoring also intact? An interesting avenue for future research would be to investigate what divers do at different depths when they (correctly) judge that some information is unlikely to be remembered. An unfortunate possibility is that although the accuracy of this metacognitive judgment is as good at deep depths as it is shallow depths, the ability or willingness of divers to implement control procedures is not.

Our results also showed that divers were aware of their impairment as measured by the V-A ratings, which is contrary to existing diver training (e.g., Shreeves, 2008; Strauss &

Aksenov, 2004). Student divers are typically told that narcosis in the 30–40 m/4–5 ata depth range will affect their judgment of performance, potentially making them unaware of narcotic impairments, negatively affecting decision making and thus their ability to respond to danger and potential problems underwater. The V-A ratings clearly showed that in both experiments, divers were aware of their impairment at deep depths and that self-assessment was not impaired. Indeed, the pattern of data suggests that loss of ability to self-assess may not be a symptom of narcosis at all, at least at depths up to 42 m/5.2 ata.

CONCLUSION

In summary, the current research contributes to our current knowledge about both memory and metacognitive performance in divers under inert gas narcosis. Indeed, it is the first study to compare memory impairment against a metacognitive index. The results showed that although both free- and cued-recall performance are impaired in these conditions, resolution is not. These results could have important implications for the kinds of control measures that are put in place to avoid diving accidents at deep depths. They suggest that although cognitive functioning is impaired, divers are aware of this impairment, which means that they can potentially compensate for it. Whether or not the impairment extends to *control*—for example, divers' ability to implement safety measures when "narked" given their knowledge of impairment—is a matter for future research. For now, at least, we can take comfort in the fact that their monitoring effectiveness is spared.

ACKNOWLEDGMENTS

This research was funded by the PADI (Professional Association of Diving Instructors) Foundation. The authors are grateful for the invaluable help in collecting this data from Nautilus Watersports, Port Vila, Vanuatu.

KEY POINTS

- Inert gas narcosis causes a spectrum of cognitive impairments in undersea divers, including a decrement in memory ability. Divers may not be able to accurately assess their level of impairment, affecting work performance and safety.
- In two underwater field experiments free- and cued-recall memory ability and ratings of self-assessed impairment were reduced at depths of 33 m to 42m, under the effects of narcosis, compared to a shallow water control. In contrast, resolution remained unaffected by depth.
- The dissociation of memory accuracy and resolution, coupled with a shift in a self-assessment of impairment, indicated that divers were able to accurately judge their decrease in memory performance under inert gas narcosis.
- This suggests that divers are better equipped to manage narcosis than previously believed.

REFERENCES

Alm, H., & Nilsson, L. (1994). Changes in driver behaviour as a function of hands-free mobile phones—A simulator study. *Accident Analysis Prevention*, *26*, 441–451.

Baddeley, A. D, De Figueredo, J. W, Hawkswell-Curtis, J. W., & Williams, A. N. (1968). Nitrogen narcosis and underwater performance. *Ergonomics*, *11*, 157–164. doi:10.1080/00140136808930952.

Bennett, P. B., & Rostain, J. C. (2002). Inert gas narcosis. In P. B. Bennett & E. H. Elliot (Eds.), *Bennett and Elliot's physiology and medicine of diving* (pp. 300–322). London: Saunders.

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied*, *8*, 44–56. doi:10.1037/1076-898X.8.1.44.

Craik, F. I. M. (1994). Memory changes in normal aging. *Current Directions in Psychological Science*, *3*, 155–158. doi:10.1111/1467-8721.ep10770653

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 474–479. doi:10.1037/0278-7393.13.3.474

Curran, V., & Hildebrandt, M. (1999). Dissociative effects of alcohol on recollective experience. *Consciousness and Cognition*, *8*, 497–509. doi:10.1006/ccog.1999.0417

De Gorordo, A., Vallejo-Manzur, F., Chanin, K., & Varon, J. (2003). Diving emergencies. *Resuscitation*, *59*, 171–180. doi:10.1016/S0300-9572(03)00236-3.

Dorrian, J., Roach, G., Fletcher, A., & Dawson, D. (2007). Simulated train driving: Fatigue, self-awareness and cognitive disengagement. *Applied Ergonomics*, *38*, 155–166.

Edmonds, C., & Douglas, W. (1990). Scuba diving fatalities in Australia and New Zealand. *Journal of South Pacific Underwater Medicine Society*, *21*(1), 2–4.

Fowler, B. (1973). Effect of hyperbaric air on short-term and long-term memory. *Aerospace Medicine*, *44*, 1017–1022.

Fowler, B., White, P. L., Wright, G. R., & Ackles, K. N. (1980). Narcotic effects of nitrous oxide and compressed air on memory and auditory perception. *Undersea Biomedical Research*, *7*(1), 35–60.

Godden, D. R., & Baddeley, A. D. (1975). Context dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*, 325–331.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*, 732–764. doi:10.2307/2281536.

Green, D., & Moses, F. (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin*, *66*, 228–234. doi:10.1037/h0023645.

Gronning, M., & Aarli, J.A. (2011). Neurological effects of deep diving. *Journal of Neurological Sciences*, *304*, 17–21. DOI:10.1016/j.jns.2011.01.021.

Hagberg, M., & Ornhagen, H. (2003). Incidence and risk factors for symptoms of decompression sickness among male and female dive masters and instructors—A retrospective cohort study. *Undersea and Hyperbaric Medicine*, *30*, 93–102.

Hamilton, K., Laliberte, M. F., & Fowler, B. (1995). Dissociation of the behavioural and subjective components of nitrogen narcosis and diver adaptation. *Undersea & Hyperbaric Medicine*, *22*, 41–49.

Hanczakowski, M., Zawadzka, K., Pasek, T., & Higham, P. A. (in press). Calibration of metacognitive judgments: Insights from the underconfidence-with-practice effect. *Journal of Memory and Language*.

Hart, A. J., White, S. A., Conboy, P. J., Bodiwala, G., & Quinton, D. (1999). Open water scuba diving accidents at Leicester: Five years experience. *Journal of Accident and Emergency Medicine*, *16*, 198–200. doi:10.1136/emj.16.3.198.

Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory and Cognition*, *30*, 67–80. doi:10.3758/BF03195266.

Higham, P. A. (2007). No special K! A signal-detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General*, *136*, 1–22. doi:10.1037/0096-3445.136.1.1

Higham, P. A. (2011). Accuracy discrimination and type-2 signal detection theory: Clarifications, extensions, and an analysis of bias. In P. A. Higham & J. P. Leboe (Eds.), *Constructions of remembering and metacognition: Essays in honour of Bruce Whittlesea* (pp. 109–127). Basingstoke, UK: Palgrave MacMillan.

Higham, P. A., & Arnold, M. M. (2007). Beyond reliability and validity: The role of metacognition in psychological testing. In R. A. DeGregorio (Ed.), *New developments in psychological testing* (pp. 139–162). Hauppauge, NY: Nova Science.

Hobbs, M. (2008). Subjective and behavioural responses to nitrogen narcosis and alcohol. *Undersea and Hyperbaric Medicine*, *35*, 175–183.

Hobbs, M., & Kneller, W. (2009). Effect of nitrogen narcosis on free recall and recognition memory in open water. *Undersea and Hyperbaric Medicine*, *36*, 73–81.

Kelly, E., Darke, S., & Ross, J. (2004). A review of drug use and driving: Epidemiology, impairment, risk factors and risk perceptions. *Drug and Alcohol Review*, *23*, 319–344.

Kiessling, R. J., & Maag, C. H. (1962). Performance impairment as a function of nitrogen narcosis. *Journal of Applied Psychology*, *46*, 91–95. doi:10.1037/h0039500.

Kneller, W., Higham, P. A., & Hobbs, M. (2012). Measuring diver impairment using a novel test of manual dexterity at 35-41m/115-135ft. *Aviation, Space & Environmental Medicine*, *83*, 54–57. doi:10.3357/ASEM.3123.2012.

Kneller, W., & Hobbs, M. (2013). The levels of processing effect under nitrogen narcosis. *Undersea and Hyperbaric Medicine*, *40*, 235–241.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370. doi:10.1037//0096-3445.126.4.349

Levett, D. Z. H., & Millar, I. L. (2008). Bubble trouble: A review of diving physiology and disease. *Postgraduate Medical Journal*, *84*, 571–578. doi:10.1136/pgmj.2008.068320.

Lieberman, P., Protopapas, A., & Kanki, B. G. (1995). Speech production and cognitive deficits on Mt. Everest. *Aviation, Space and Environmental Medicine*, *66*, 857–864.

Logie, R. H., & Baddeley, A. D. (1985). Cognitive performance during simulated deep-sea diving. *Ergonomics*, *28*, 731–746. doi:10.1080/00140138508963194.

Luna, K., Higham, P. A., & Martin-Luengo, B. (2011). The regulation of memory accuracy with multiple answers: The plurality option. *Journal of Experimental Psychology: Applied*, *17*, 148–158. doi:10.1037/a0023276.

Manzey, D. (2000). Monitoring of mental performance during spaceflight. *Aviation, Space, and Environmental Medicine*, *71*(9), A69–75.

Manzey, D., Lorenz, B., Heuer, B., & Sangals, J. (2000). Impairments of manual tracking performance during spaceflight: More converging evidence from a 20-day space mission. *Ergonomics*, *43*, 589–609.

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 509–527. doi:10.1037/a0014876.

Morrison, J. B., & Zander, J. K. (2008). *The effect of pressure and time on information recall* (Contract report). Taree, Australia: Sheerwater Engineering.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*, 109–133. doi:10.1037/0033-2909.95.1.109

Nelson, T. O., Dunlosky, J., White, D. M., Steinberg, J., Townes, B., & Anderson, T. (1990). Cognition and metacognition at extreme altitudes on Mount Everest. *Journal of Experimental Psychology: General*, *119*, 367–374.

Nelson, T. O., McSpadden, M., Fromme, K., & Marlatt, G. A. (1986). Effects of alcohol intoxication on metamemory and on retrieval from long-term memory. *Journal of Experimental Psychology: General*, *115*, 247–254.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 1–45). New York, NY: Academic Press.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA: MIT Press.

Philp, R. B., Fields, G. N., & Roberts, W. A. (1989). Memory deficit caused by compressed air equivalent to 36 metres of seawater. *Journal of Applied Psychology*, *74*, 443–446. doi:10.1037/0021-9010.74.3.443.

Pollack, I., Norman, D., & Galanter, E. (1964). An efficient nonparametric analysis of recognition memory. *Psychonomic Science*, *1*, 327–328. doi:10.1037/h0031246.

Pulak, C. (1998). The Uluburun shipwreck: An overview. *The International Journal of Nautical Archaeology*, *27*, 188–224. doi:10.1111/j.1095-9270.1998.tb00803.x

Ross, H. E. (1989). Perceptual and motor skills of divers underwater. *International Review of Ergonomics*, *2*, 155–181.

Rotello, C. M., Masson, M. E. J., & Verde, M. F. (2008). Type 1 error rates and power analyses for single-point sensitivity measures. *Perception & Psychophysics*, *70*, 389–401. doi:10.3758/PP.70.2.389.

Shreeves, K. (2008). The diver within: The physiology of diving. In D. Richardson, J. Kinsella, & K. Shreeves (Eds.), *The encyclopedia of recreational diving* (3rd ed., pp. 5.20–5.23). Santa Ana, CA: PADI.

Strauss, M., & Aksenov, I. (2004). *Diving science*. Leeds, UK: Human Kinetics.

Tetzlaff, K., Leplow, B., Deistler, I., Ramm, G., Fehm-Wolfsdorf, G., Warninghoff, V., & Bettinghausen, E. (1998). Memory deficits at 0.6 MPa ambient air pressure. *Undersea and Hyperbaric Medicine*, *25*, 161–166.

Van Rees Vellinga, T. P., Verhoeven, A. C., Van Dijik, F. J. H., & Sterk, W. (2006). Health and efficiency in trimix versus air breathing in compressed air workers. *Undersea and Hyperbaric Medicine*, *33*, 419–427.

Malcolm Hobbs, of the University of Winchester, received his PhD in experimental psychology from the University of Southampton in 2004.

Phillip A. Higham, of the University of Southampton, received his PhD in cognitive psychology from McMaster University, Canada, in 1993.

Wendy Kneller, of the University of Winchester, received her PhD in experimental psychology from the University of Southampton in 2002.

Date received: May 23, 2013
Date accepted: October 3, 2013