
Visualizing Uncertainty: The Impact on Performance

Susan S. Kirschenbaum, J. Gregory Trafton, Christian D. Schunn and Susan B. Trickett

Human Factors: The Journal of the Human Factors and Ergonomics Society 2014 56: 509 originally published online 30 July 2013
DOI: 10.1177/0018720813498093

Visualizing Uncertainty: The Impact on Performance

Susan S. Kirschenbaum, Naval Undersea Warfare Center, Newport, Rhode Island, **J. Gregory Trafton**, Naval Research Laboratory, Washington, DC, **Christian D. Schunn**, University of Pittsburgh, Pittsburgh, Pennsylvania, and **Susan B. Trickett**, Naval Research Laboratory, Washington, DC

Objective: This work investigated the impact of uncertainty representation on performance in a complex authentic visualization task, submarine localization.

Background: Because passive sonar does not provide unique course, speed, and range information on a contact, the submarine operates under significant uncertainty. There are many algorithms designed to address this problem, but all are subject to uncertainty. The extent of this solution uncertainty can be expressed in several ways, including a table of locations (course, speed, range) or a graphical area of uncertainty.

Method: To test the hypothesis that the representation of uncertainty that more closely matches the experts' preferred representation of the problem would better support performance, even for the nonexpert., performance data were collected using displays that were either stripped of the spatial or the tabular representation.

Results: Performance was more accurate when uncertainty was displayed spatially. This effect was only significant for the nonexperts for whom the spatial displays supported almost expert-like performance. This effect appears to be due to reduced mental effort.

Conclusion: These results suggest that when the representation of uncertainty for this spatial task better matches the expert's preferred representation of the problem even a nonexpert can show expert-like performance.

Application: These results could apply to any domain where performance requires working with highly uncertain information.

Keywords: decision making, naturalistic decision making, cognitive processes, knowledge representation, problem solving, reasoning

INTRODUCTION

This has been called the "information age" and people expect to use information to help them make decisions. However, not everything that poses as information is accurate and/or complete. Thus, the knowledgeable person must assess the underlying uncertainty of the information used. Sometimes the information comes with some indicator of its reliability (e.g., 50% chance of rain) and sometimes it does not (a high of 68º). The indicator can be numeric, as the previous example, or semantic (e.g., very likely), or graphical (e.g., variance bars on a graph). This paper examines the ways in which the external representation of uncertainty impacts the interpretation of information and hence task performance. That uncertainty information impacts decision making should not be a question; however, the research results are inconclusive, perhaps due to the format of the uncertainty information (Gigerenzer & Brighton, 2009) or the training of the decision maker.

There are many variants of uncertainty. The term could refer to noise in the information, statistical variability, nondeterministic relationship between action and consequences, or even the psychological reaction to difficult problems. In this paper we shall limit ourselves to any factor that makes the information less than 100% accurate. We will provide more detail on the specific components when we discuss the domain.

Although there are numerous papers offering suggestions for how to depict uncertainty (Pang, Wittenbrink, & Lodha, 1997) and some that apply usability engineering to their displays (Slocum, Cliburn, Feddema, & Miller, 2003), there are few that link the form of the visualization to human performance. For example Kirschenbaum and Arruda (1994) showed that compared to a textual indicator, a visual display, referred to as an area of uncertainty (AOU) ellipse, significantly improved accuracy in a

---

Address correspondence to Susan S. Kirschenbaum, Naval Undersea Warfare Center, 1176 Howell St., Code 2501, Building 1171/3, Newport, RI 02881, USA; e-mail: susan.kirschenbaum@navy.mil.

HUMAN FACTORS
Vol. 56, No. 3, May 2014, pp. 509–520
DOI: 10.1177/0018720813498093

submarine location task. Finger and Bisantz (2002) found similar improvements when using a degraded or blended visual icon for a radar identification task. Both of these studies were spatial tasks and the spatial form of the uncertainty presentation matched that of the basic form of the task.

In contrast, Mahan and colleagues (Mahan, Kirschenbaum, Jilg, & Marino, 1998) used a temporal task, the arrival time of a ship at a pier, and varied amount of uncertainty and presentation mode. In the dynamic (temporal) condition, uncertainty was represented by an animation in which a larger/faster animation effect indicated greater uncertainty. Again, under the most difficult conditions, the display that more closely matched the associated variable and the representation of the problem supported significantly better performance. In this case, a dynamic (temporally animated) display improved performance on this temporal task while spatial and numeric displays of uncertainty did not.

Representation Match Hypothesis

Following Zhang's (1997) definitions, "external representations are defined as knowledge and structure in the environment . . . and as external rules, constraints, or relations embedded in physical configurations. . . . In contrast, internal representations are the knowledge and structure in memory" (p. 180). Internal representations can be made external as when one draws a figure or writes an equation to represent the problem or task. Conversely, an external representation is converted into an internal one as one learns about the problem domain. Interestingly, there is evidence from neuroimaging data to indicate that mental imagery and perception of external visual images display similar brain activation patterns (Borst & Kosslyn, 2008; Kosslyn, Ball, & Reiser, 1978; William, Scott, Marie, & Stephen, 2009).

It is our conjecture that external representations of uncertainty that more closely match the internal representation used by the problem solver will better support decision making. However, as we have no access to the internal representation, we propose an alternative that does not rely on the internal representation. Specifically, with respect to uncertainty, the findings cited previously suggest that performance in a domain should be better (i.e., more accurate or faster) when the uncertainty is represented externally in the same general format (spatial, temporal, etc.) as the key variable(s) in the problem. Thus, if the problem is fundamentally temporal (e.g., predict when will something occur), then the uncertainty representation should capture the time dimension; if it is fundamentally computational, then it should be represented by numerals; if it is fundamentally spatial (e.g., judge where something is), then the representation of uncertainty should be spatial. This does not mean that the problem solver's representation is an exact copy (or even diagram) of the real world, but that it captures the elements of the problem that are required for problem solving in a way that is congruent with the problem. Thus, a timeline can be a temporal display if appropriate to the task. Note that even if the domain is generally spatial, the particular problem might not be. For example, the problem might be to compute the distance between two points in space rather than determine the location of an object. In that case, a numeric representation might be better. In the case of the problem presented here, the task is to judge if the location of an object is sufficiently well known to take action, and thus a spatial representation is predicted to be more useful.

Representation match per se does not tell the entire story. Why should matching the representation of uncertainty to the general form of the problem representation improve performance? One reason might be that external congruent representation reduces the internal mental work of translating from one kind of representation (e.g., numeric) to another (e.g., spatial) representation.

The external representation might not be limited to the visible/audible elements of the problem, but could also include other information such as uncertainty or directional/force vectors. Experts use this information to make decisions and solve problems, and experts include them explicitly in their problem representations. For example, Larkin and Simon (1987) found that expert physicists included force vectors in their representations of simple machines. This additional information was incorporated into the drawings of the visible elements of the problems (pulleys, ramps, etc.). It was typical of the drawings of experts but not student-novices. This might be because adding the vector required more mental effort than the students could muster. Kirschenbaum (1992) also found evidence for experience differences creating an accurate external representation when novice, journeymen, and expert submariners transformed numeric course, speed, and bearing into traditional nautical line-of-sight vector drawings. These drawings were created from memory and without any intermediary work on paper. Thus, the transformation from numeric to spatial representation required internal cognitive work. In both studies, the experts were more accurate than novices.

Perhaps, this performance difference between experts and novices performance could be partially alleviated by augmenting external representations for novices or aiding them in the creation of the external representation. Nadav-Greenberg and Joslyn (2009) found that a representation of uncertainty improved nonexpert decision making for a task using weather information. An earlier paper reported improvement for both novices and professional forecasters (Nadav-Greenberg, Joslyn, & Taing, 2008).

Objective measurement of mental effort is difficult, so we took the approach of analyzing speech utterances for indicators of mental effort in the form of changing or adding to the explicit information that was given. For example, the comments “The TMA solution range is about what I expect from my mental calculation” or “I think his range is closer, and he’s headed reciprocal to what the system solution is” clearly indicate mental effort to confirm or disconfirm the computer. These are well-practiced skills for expert submariners who engage in what they call “mental gym” to develop speed and accuracy.

From here forward we limit our discussion to the representation of uncertainty. One challenge to studying the underlying nature of key variables in a problem involves determining how the experienced performer prefers to represent the problem. Our first approximation to an answer for this question was to analyze the problem in the “real” world and look for logical representations. Thus, if the problem is a spatial one (e.g., Kirschenbaum & Arruda, 1994), one might expect a spatial representation of uncertainty. If, in contrast, the problem is a temporal one (Mahan et al., 1998), one could predict that the uncertainty representation would be temporal. Another way to investigate this idea is to examine the form of the information that the experienced person uses when he or she has several formats available. By capturing gaze frequency and duration, the eye-tracker can help with this examination. In a pilot study of five highly experienced submarine officers (all prior commanding or executive officers) we did just that. The eye-tracking results showed that these experts spent 65.8% (*SD* = 25.4%) of the time looking at spatial representations of the problem as compared to 26.7% (*SD* = 30.1%) of the time looking at either tabular or textual information. Although not statistically significant due to the large variance and small sample, that is a difference of more than two to one.

In the submarine domain there are typically several different explicit representations of statistical uncertainty. If, however, the representation of uncertainty is not congruent with the preferred problem representation (e.g., a numeric or textual representation of a spatial problem), the problem solver must mentally add uncertainty to the perceptual picture, transformed from the numeric or textual values. This could result in longer time to solve the problem or a less accurate solution. Thus, placing the external representation of uncertainty on the same display as the representation of the spatial relationships between the contact and own ship reduces the cognitive effort required to assess the goodness of the solution and facilitates performance. In summary, representations that require additional mental effort may impede performance, either in terms of the time required, or accuracy, or both.

The remainder of this paper will first describe key features of the submarine domain and then address data from the eye-tracker that points to most useful information format. Finally, we will discuss the central experiment in which uncertainty representation format was manipulated.

Domain Background

Before discussing the experiments, the reader needs to understand a bit about the submarine domain to understand the visualizations being examined and the nature of uncertainty that is the focus of problem solving. The task of determining the range, course, and speed of underwater contacts in a passive sonar scenario is called submarine target motion analysis (TMA). Locating the contact is a team sport with specific responsibilities assigned to specific individuals. This requires an enlisted sonar technician to identify the contact in the noise of the undersea environment, an enlisted TMA technician to employ the various TMA techniques and algorithms to localize the contact, and an officer (officer of the deck, OOD) to maneuver the ship and decide when the "solution" is sufficiently accurate for the task at hand. The OOD does not perform TMA but does evaluate the quality (accuracy and certainty) of the solution to support decision making. The focus of this study was the OOD's use of uncertainty displays to support his decision making. (At the time of the study all submariners were men.) The enlisted sonar tech and TMA tech were simulated. As on a Navy submarine, the OOD was able to look at their displays, (virtually) over their shoulders and thus observe the displays of uncertainty. The OOD does not typically have his own dedicated displays.

Localization is a particularly difficult problem, in part because of the physics of sound transmission through water (Urick, 1983). Sound reflects, refracts, and scatters due to pressure, temperature, salinity, and turbidity. The information "seen" is rather like a reflection in a fun-house mirror. Thus, one source of the uncertainty is the noisy nature of the basic data, the received bearing to the contact.

The U.S. submarine force primarily uses passive sonar, that is, they just listen. When something makes noise underwater that noise is picked up by hydrophones, processed, and displayed to the sonarman as a swath of coherent noise against a background of random noise (see Figure 1). What he sees tells him that there is some noise source roughly at a given bearing.

Unlike a visual contact, range, speed, and direction of movement, or even the identity of the contact, are not observable. Because these features of the contact are not directly observable, there is uncertainty in the range, speed, and course of the contact, but because there are limits to how fast an undersea contact is likely to be going, the greatest uncertainty is in range.

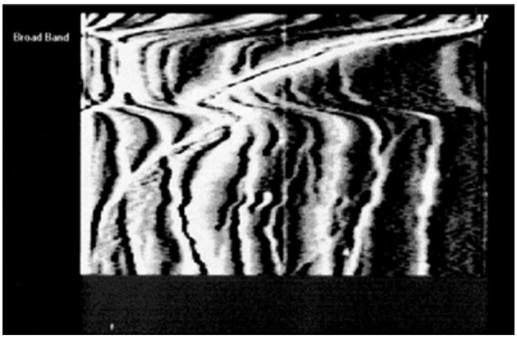


Figure 1. Example of a sonar display.

Not only are the data noisy, but solving the problem is mathematically underdetermined with two known quantities (bearing and time) and three unknowns (course, speed, range). There are numerous techniques and algorithms to determine the solution using data collected over time (Ferkinhoff, Nguyen, Hammel, & Gong, 1993), but it is frequently difficult or impossible to determine which one will work best in any given situation. Lastly, the solution techniques and/or algorithms contain their own sources of uncertainty, either statistical or procedural. These complications are always considerations in solving the TMA problem. Thus, determining the location of a sound source ("contact") is a very difficult problem because the information is uncertain, there is much at risk, and the situation is always dynamic. Given enough time, the solutions usually converge, but because this is a dynamic problem, speed and accuracy have a U-shaped relationship. It takes some time (and maneuvers) to develop a good solution, but taking too long might allow the other platform to detect own ship, or it might just drive out of the area.

All of the aforementioned contribute to the uncertainty associated with submarine decision making. To reiterate, the sonar data are noisy due to the way sound is transmitted underwater and the noise in the environment. Second, these data provide only bearings and the time the bearings arrive while the solution must give course, speed, and range to the sound source. Third, all of the computations (there are many possible algorithms) used to transform these noisy bearings have statistical uncertainty and, due to the underconstrained nature of the problem, result in multiple possible solutions. With proper maneuvering, the solutions do converge over time, but waiting too long causes its own problems. Thus, there are multiple sources for the uncertainty.

Following the representation match hypothesis, we predicted that when only one type of display was provided, spatial displays of uncertainty would lead to better performance than tabular ones. To test this prediction we manipulated the display of locational uncertainty while maintaining the single best point solution. Uncertainty was displayed in either tabular or spatial format, but not both. This allowed us to determine the effect of uncertainty display format on performance. Because the spatial uncertainty displays did not require mentally adding uncertainty to the picture, we predicted fewer spatial transformations would be required to determine a solution.

METHOD

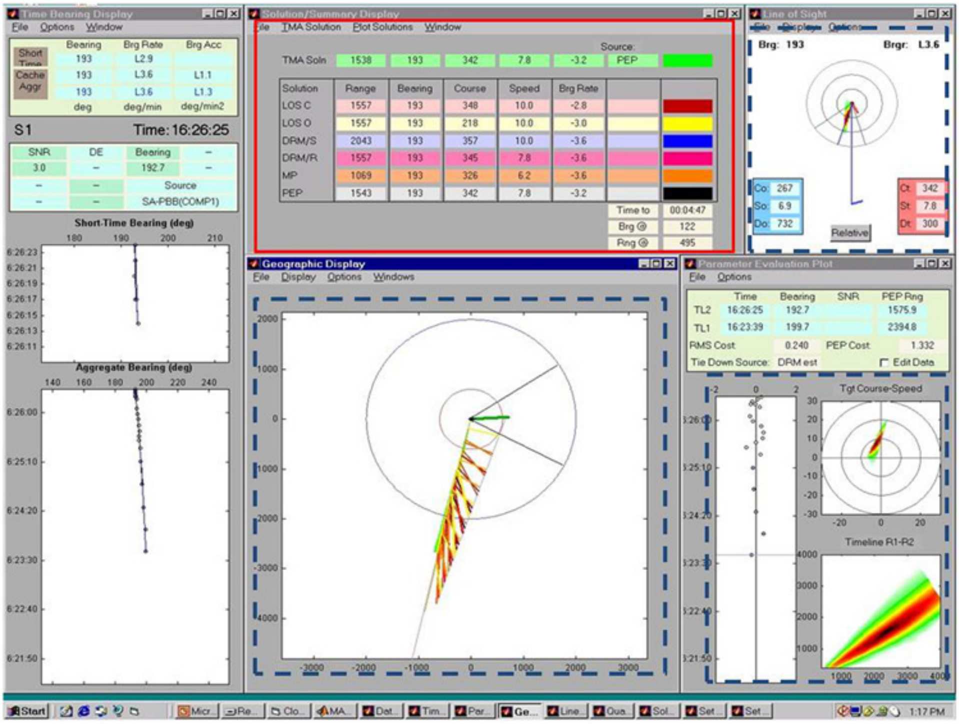
Participants

There were 16 submarine officers who participated. One was eliminated because he was called away and was unable to complete the experiment. Although they all held the rank of lieutenant and all were qualified as officer of the deck, they varied in expertise due to their specific shipboard experience. All participants were instructors at the U.S. Submarine School and, as such, were highly qualified in the subjects that they taught. Participants had a mean of 9 (*SD* = 4) years in the Navy ranging from 5 to 14 years of Naval experience. As the participants were all instructors, expertise was defined according to what they taught. Experts taught the skills needed for this task, nonexperts taught other subjects. For example, the experts taught tactics (i.e., how to respond to a hostile contact that necessarily includes treatment of uncertainty in sonar) while nonexperts taught another subject such as navigation (i.e., how to use a maritime chart, rules of the road, which does not involve analysis of passive sonar data). Seven experts and eight nonexperts completed the experiment.

Apparatus

Combat system toolbox (CS Toolbox). Submariners used a simulated combat system called the CS Toolbox that contains an unclassified version of their usual tools. This system is written in MatLab™ and was originally designed to test new algorithms. It was run in a classroom on a laptop computer and displayed on a 17-inch monitor. A scan converter was connected to the monitor to capture the display onto videotape.

The participants were playing the role of a decision maker (i.e., OOD), not an operator (i.e., sonar or TMA tech). Therefore, the solutions were displayed as if they came from operators or algorithms. This arrangement mimics shipboard conditions where the decision makers do not actually interact with equipment, although they do look at the screens and solutions generated by operators. Participants were able to make maneuvers as they would aboard ship. The CS Toolbox kept track of both generated solutions and the scenario truth. Uncertainty was displayed in a number of ways (see Figure 2), including a table of solutions, a spatial area of uncertainty, and a set of solution lines colored to match the regions and solutions in the table and area of uncertainty. (This coloring was arbitrary and intended to indicate what solutions went together, not their likelihood.) In Figure 2, the spatial representations are enclosed in dashed lines and the table of solutions is enclosed in a solid line. In order to separate the two classes of uncertainty display for the experiment, two separate displays were developed, one with only a tabular representation (see Figure 3a) and one with only spatial representations of uncertainty (Figure 3b), providing two relatively clean experimental conditions with high external validity in the complex domain. These two displays are approximately equivalent in that both include the essential information of bearing, course, speed, and range and can be interpreted to indicate the extent of the solution disagreement and both give the most likely solutions from the same set of algorithms. Where they are not equivalent is that the graphical representation, especially the parameter estimation plot, considers all possible values of the solutions and grades them for agreement. Thus, the spatial displayed did provide nominally more



	Bearing	Brg Rate	Brg Acc
Short Time	193	L2.9	
Cache Aggr	193	L3.6	L1.1
	193	L3.6	L1.3
	deg	deg/min	deg/min2

S1 Time: 16:26:25

SNR	DE	Bearing	–
3.0	–	192.7	–
–	–	Source	
–	–	SA-PBB(COMP1)	

Source: PEP

Solution	Range	Bearing	Course	Speed	Brg Rate
TMA Soln	1538	193	342	7.8	-3.2
LOS C	1557	193	348	10.0	-2.8
LOS O	1557	193	218	10.0	-3.0
DRM/S	2043	193	357	10.0	-3.6
DRM/R	1557	193	345	7.8	-3.6
MP	1069	193	326	6.2	-3.6
PEP	1543	193	342	7.8	-3.2

Time to 00:04:47
Brg @ 122
Rng @ 495

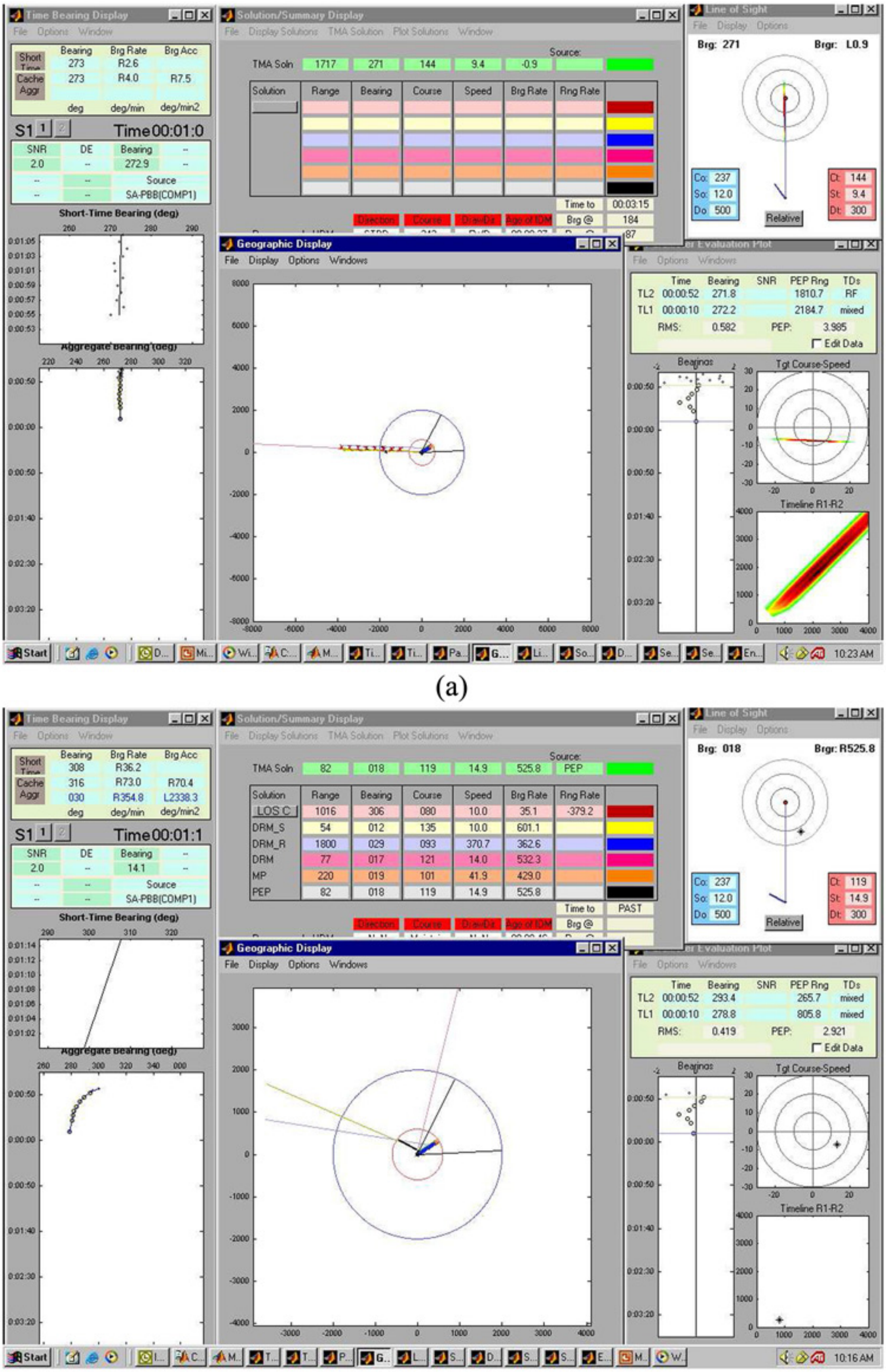
Brg: 193 Brgr: L3.6
Co: 267
So: 6.9
Do: 732
Ct: 342
St: 7.8
Dt: 300

	Time	Bearing	SNR	PEP Rng
TL2	16:26:25	192.7		1575.9
TL1	16:23:39	199.7		2394.8

RMS Cost 0.240 PEP Cost 1.332
Tie Down Source: DRM est ☐ Edit Data

Figure 2. This figure shows the combat system (CS) Toolbox. Uncertainty is displayed as both a spatial set (enclosed in a dashed line) of possible solutions and a table of solutions (enclosed in a solid line) to the localization problem. The center window is a geospatial (GEO) or bird's eye display, with own ship in the center and bearing lines at the current time and in the past extending from own ship. The most likely possible solutions are the lines between the two bearing lines. Moving clockwise from the left, the other windows show the following. The top left window shows a time-bearing display of measured sonar data at two different aggregations with the data plotted below the table. The next window shows a table of solutions displaying the numeric values of some possible solutions with the recommended solution at the top. (Notice that the greatest variation is in range, the most difficult component of localization to determine. Solutions below the "TMA Solin" line are in no particular order.) The next window is a traditional maritime line of sight (LOS) display with own ship at the bottom and vectors representing bearing and possible courses and speeds. Below that is a parameter evaluation plot (PEP) and associated input data. The PEP (bottom right) evaluates the likelihood of the full set of solutions at time one and time two. To the left of it is a bearing scatter graph.

information, not just a different modality of information. While there is more screen real estate devoted to the spatial representation, the data tables are important inputs to some of the mental calculations that submariners are trained to perform to validate any computerized solutions. No table could show all possible solutions, so the table shows only the most likely for each solution. Furthermore, because there is always uncertainty in this problem and because uncertainty has been communicated to submariners by the system for at least the past 25 years, there was no condition that did not use some means of displaying uncertainty. Such a condition would have been impossible from their perspective. As this experiment used real submariners as the subject population, we were limited in how we could manipulate the displays and still maintain an information environment in which their expertise was valid.



(a)

Time Bearing Display

	Bearing	Brg Rate	Brg Acc
Short Time	273	R2.6	
Cache Aggr	273	R4.0	R7.5
	deg	deg/min	deg/min2

S1 Time 00:01:0

SNR	DE	Bearing	--
2.0	--	272.9	--
--	--	Source	
--	--	SA-PBB(COMP1)	

Solution/Summary Display

TMA Soln	1717	271	144	9.4	-0.9	Source:

Solution	Range	Bearing	Course	Speed	Brg Rate	Rng Rate	
							🟥
							🟨
							🟦
							🟪
							🟧
							⬛

Time to 00:03:15

Brg @ 184

87

Line of Sight

Brg: 271 Brgr: L0.9

Co: 237 So: 12.0 Do: 500

Ct: 144 St: 9.4 Dt: 300

Parameter Evaluation Plot

	Time	Bearing	SNR	PEP Rng	TDs
TL2	00:00:52	271.8		1810.7	RF
TL1	00:00:10	272.2		2184.7	mixed

RMS: 0.582 PEP: 3.985

☐ Edit Data

10:23 AM

(b)

Time Bearing Display

	Bearing	Brg Rate	Brg Acc
Short Time	308	R36.2	
Cache Aggr	316	R73.0	R70.4
	030	R354.8	L2338.3
	deg	deg/min	deg/min2

S1 Time 00:01:1

SNR	DE	Bearing	--
2.0	--	14.1	--
--	--	Source	
--	--	SA-PBB(COMP1)	

Solution/Summary Display

TMA Soln	82	018	119	14.9	525.8	Source: PEP

Solution	Range	Bearing	Course	Speed	Brg Rate	Rng Rate	
LOS C	1016	306	080	10.0	35.1	-379.2	🟥
DRM_S	54	012	135	10.0	601.1		🟨
DRM_R	1800	029	093	370.7	362.6		🟦
DRM	77	017	121	14.0	532.3		🟪
MP	220	019	101	41.9	429.0		🟧
PEP	82	018	119	14.9	525.8		⬛

Time to PAST

Brg @

Line of Sight

Brg: 018 Brgr: R525.8

Co: 237 So: 12.0 Do: 500

Ct: 119 St: 14.9 Dt: 300

Parameter Evaluation Plot

	Time	Bearing	SNR	PEP Rng	TDs
TL2	00:00:52	293.4		265.7	mixed
TL1	00:00:10	278.8		805.8	mixed

RMS: 0.419 PEP: 2.921

☐ Edit Data

10:16 AM

Figure 3. The separate displays with only (a) spatial uncertainty or only (b) tabular uncertainty. These are essentially the same displays as Figure 2, but with the alternative solutions removed from either the table of solutions or the line-of-sight, parameter evaluation plot (PEP), and geographic windows.

ASL Model 501 eye-tracker. In order to track what information was being used, an ASL Model 501 eye-tracker was also used for most of the participants. The eye-tracker was mounted on a chinrest to stabilize the head for more accurate tracking. Difficulties tracking five of the participants limited use of the eye-tracker, generally due to the Navy-issued wire rim eyeglasses that produce significant reflections. Of those who could be tracked, four were classified as experts and five were nonexperts.

A lapel microphone connected to a Sony mini digital camcorder mounted on a tripod recorded the general scene and participant verbal protocols.

Scenarios

The scenarios were relatively simple, with only one contact. Two scenario types (geometry/range) were created, *a* and *b*. To reduce the possibility of participants sharing their experiences with colleagues, two versions of each scenario were created by rotating the geometry by a random amount, for a total of four scenarios, *a1/2* and *b1/2*. Each participant was given one of the *a* scenarios and one of the *b* scenarios, with order counterbalanced across participants.

Task

The task of the participants was to determine when there was an adequate solution on the contact to shoot a torpedo at it. This was a decision-making task, as the system used an automated "operator" to actually manipulate solutions. This arrangement is typical of many Naval decision tasks. This was not just a waiting task, but required maneuvering the boat to facilitate improving the solution. The geometry of the problem (relative positions of own ship and the other platform) determines what constitutes a good maneuver and how many maneuvers are required to sufficiently refine the solution. Because the torpedo has its own sonar, the goal was to determine when the solution was adequate to hit the target and not to work the problem beyond that point.

Procedure

After signing an informed consent form, participants were introduced to the CS Toolbox. Although participants were familiar with the components of the toolbox, none had used the toolbox itself prior to the experiment. Initial training was with the full CS Toolbox. Reading from a script, the experimenter gave the participants a guided tour of the CS Toolbox so that they would know how and where to find their accustomed tools. The tour began with the upper left-most window and proceeded clockwise. They were also taught how to use some experiment-specific features such as how to change own ship's course and speed by dragging a speed-course vector line to the desired setting. (Shipboard, this would be done by voice command to the helm.) As every submarine is different, having different versions of the contact management software, all questions could not be anticipated in advance so the demonstration was adapted to answer individual participants' questions. To assure their comfort with the Toolbox, they also worked one sample problem before beginning the experiment.

During the experiment two different displays were used (see Figure 3), one with uncertainty only represented spatially and one with only a tabular representation of uncertainty. All participants solved one scenario with each version, at a single session. After the practice scenario, they were introduced to the first test system (tabular or spatial) and the first scenario was run. Subsequently, the second display variant was introduced and the second scenario was run. The presentation order and scenarios were counterbalanced.

Following the demonstration and sample scenario, participants were asked to look at the display and follow directions while an experimenter calibrated the eye-tracker.

Instructions. The participants were instructed that their task was to determine when an adequate solution had been achieved to successfully shoot a torpedo at the contact. Participants were instructed to tell the experimenter when they had a "firing solution." The problem was stopped at that point because the simulation did not include weapons. While they worked, participants were videotaped and asked to talk aloud. Following the experiment, all participants then answered biographical questions and were debriefed. Each experiment took approximately 2 hours.

ANALYSIS AND RESULTS

Two performance metrics were analyzed, accuracy and speed, to come to a solution. Better performance can be defined as faster solutions or more accurate solutions. The submarine world is not necessarily subject to a speed/accuracy tradeoff. Sometimes not acting when there is sufficient information reduces accuracy because both the contact and own ship are moving. The movement may make the solution less accurate due to changes occurring after the last observation or good solution such as, for example, a maneuvering contact, or a contact moving out of sensor/weapon range. Moreover, localization for weapon placement need not be precise because the torpedo has its own homing sonar. Thus, waiting for an overaccurate solution can be counterproductive and may endanger own ship by increasing the probability of counterdetection. For these reasons, while retaining the timeliness measure, we opted for a binary accuracy criterion of weapon hit or miss, rather than a continuous degree of error measure. This is the standard to which Naval officers are trained. Calculation of hit or miss was based on standard U.S. Navy methods.

Accuracy

Overall, participants had more hits than misses (27 hits vs. 10 misses), however, they were not evenly distributed across conditions with 0.80 probability of a hit for the spatial condition only and 0.62 probability of a hit for the tabular condition. These condition differences were statistically significant, $\chi^2(1) = 6.5$ $p < .05$.

Expertise effects. Overall, experts were significantly more likely to have a hit than nonexperts, $\chi^2(1) = 3.82$, $p < .02$, with the probability of a hit for an expert at 0.84 and for a nonexpert at 0.54. There was also an expertise by display interaction (see Figure 4). Display did not matter for experts, but the nonexperts had significantly more hits with the spatial display of uncertainty, $\chi^2(1) = 8.5$, $p < .01$. Notice that with the spatial display, the nonexperts' hit rate (0.75) nearly matched that of the experts (0.84).

Time to Complete

Time to complete the task was similar, regardless of condition (spatial $M = 582$ seconds, $SEM = 59$; tabular $M = 695$ seconds, $SEM = 71$) with no significant differences. While there appeared to be interaction effects, none reached the level of significance, probably due to large variances. Importantly, the condition with worse accuracy seemed to show a nonsignificant, but suggestive, trend toward longer solution time, so there were no complicating speed-accuracy tradeoffs across conditions.

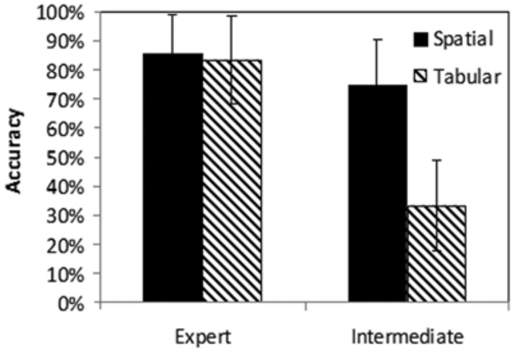


Figure 4. Accuracy as a function of expertise and condition.

Eye Tracking

The eye tracking data showed that there were few differences and none of them statistically significant due to large variances. In both conditions participants spent over 30% of the time looking at the principal spatial display, the Geo (the large central geositational display with own ship in the center, Figure 2). The one suggestive difference between the two conditions was that in the tabular condition they appeared to spend more time looking at the table (Figure 5), but again, this was not a large difference (25% vs. 15%) nor was it statistically significant.

Mental Effort

The verbal protocols were transcribed and all utterances coded for indications of mental effort as defined by mention of conclusions or projections that could not be read directly from the information on the display. As the participants had no paper or other means of externally recording or noting ideas, such utterances likely represented mental efforts rather than summaries of external problem solving. A quarter

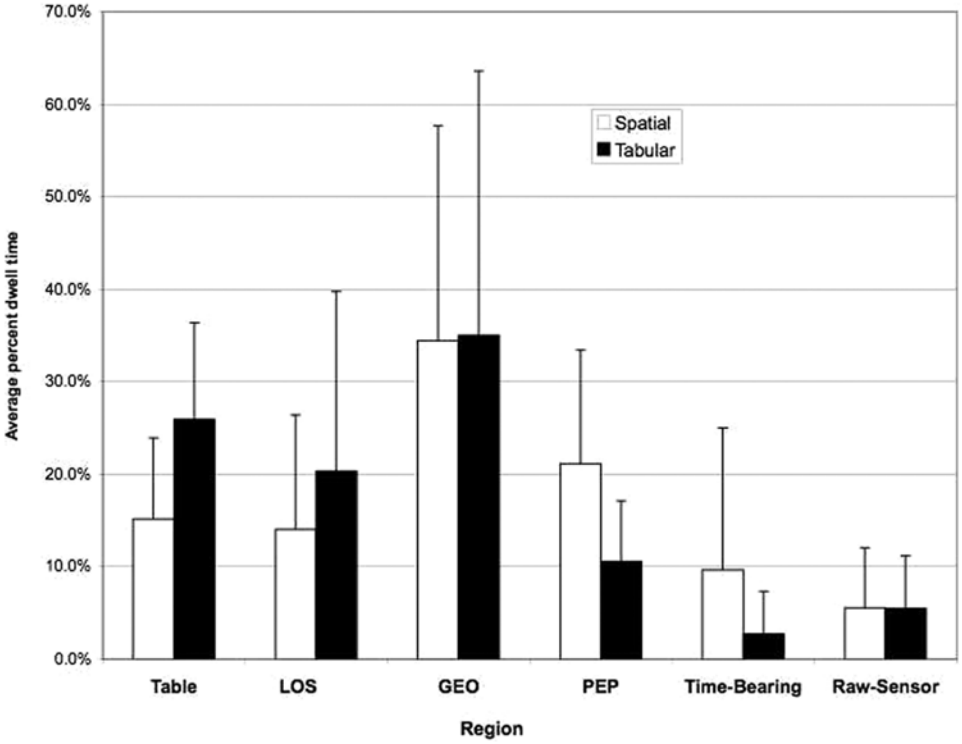


Figure 5. Mean saccade frequency by region. Category names correspond to areas in Figure 2, including Table (of solutions), line of sight (LOS), geospatial (GEO), parameter evaluation plot (PEP), Time-Bearing (side left), and raw-sensor (table top left).

of the transcripts were coded independently by two individuals, one experimenter and one undergraduate assistant. The interrater reliability kappa was 0.74, $p < .001$.

As participants' talk varied in the quantity of verbal utterance, both overall and across scenarios, the proportion of mental effort utterances rather than the raw number of utterances was used for analysis. Mean proportion of such utterances was 0.27 (*SEM* = 0.02) for the spatial condition and 0.31 (*SEM* = 0.03) for the tabular condition, representing a significantly greater proportion of such mental effort utterances when using the tabular display than when using the spatial representation, $F(1, 14) = 4.91$, $p < .05$, with a moderate effect size, Cohen's $d = .30$. This pattern held for both experts and nonexperts, however experts made slightly fewer spatial transformations overall; experts, $M = 0.26$ (*SEM* = 0.045), nonexperts, $M = 0.3$ (*SEM* = 0.03), again with a moderate effect size, Cohen's $d = .47$.

DISCUSSION

This experiment supported the representation match hypothesis in general. This general effect is not surprising as the submarine localization problem is a spatial one, and eye tracking data suggested that experts used spatial representations of uncertainty most often. Then, not surprisingly, performance on the TMA task was both more accurate and more rapid when problem solvers were given uncertainty information in spatial formats than when they were given uncertainty information in tabular formats (in the form of different TMA estimates).

The unpredicted expertise effect might be the most important result from these studies. Contrary to our expectation, performance enhancements with the spatial representation of uncertainty were largely limited to nonexperts. Experts are apparently so highly practiced that they can do the task with whatever information and displays they are given. On the other hand, the effect of representation for nonexperts shows that their performance can be improved significantly with the right representation. Perhaps we should have predicted this result, given the literature (Nadav-Greenberg & Joslyn, 2009).

These data have implications for both display design and training. The most obvious implication is, that at least for this spatial task and for nonexperts, uncertainty should be represented and that representation should be spatial. A closer read of the data could suggest that the uncertainty representation could be presented in various formats when an expert is in control, although additional studies may reveal benefits for experts of the preferred format in more complex scenarios.

A second implication is that spatial uncertainty displays could provide a kind of scaffolding that supports training. Anecdotally, we can say that experts know that any solution is uncertain and have a feeling for both how uncertain it is and how much uncertainty is acceptable. We speculate that explicitly providing a spatial representation of uncertainty, perhaps including fading the uncertainty representation as experience grows, could help the trainee to develop his knowledge in this critical area.

These experiments were limited to a single contact. With multiple contacts, as is often the case in the more crowded waters near shore and in constrained waters such as in straits and gulfs, multiple areas of uncertainty might overlap or clutter an already cluttered screen. Additional research is needed to address these situations.

ACKNOWLEDGMENTS

The authors wish to thank the programming and experimentation efforts of Megan Gibson and the officers of the U.S. Navy Submarine School, Groton, Connecticut, for their participation in this research. We also wish to thank the thoughtful comments of three anonymous reviewers. This work was supported by the Office of Naval Research under task TD2212, Astrid Schmidt-Nielsen, Science Officer.

KEY POINTS

- Uncertainty can be represented in a number of ways.
- For a spatial task such as determining the location of an object, a spatial representation was found to support more accurate and timely decision making.
- These effects were dependent on experience, with no differences for experts but strong differences for nonexperts.

REFERENCES

Borst, G., & Kosslyn, S. M. (2008). Visual mental imagery and visual perception: Structural equivalence revealed by scanning process. *Memory and Cognition*, *36*, 849–862.

Ferkinhoff, D. J., Nguyen, C. T., Hammel, S. E., & Gong, K. F. (1993). *Performance characterization of artificial neural networks for contact tracking*. Paper presented at the OCEANS '93 "Engineering in Harmony With Ocean" Proceedings, Victoria, Canada.

Finger, R., & Bisantz, A. M. (2002). Utilizing graphical formats to convey uncertainty in a decision-making task. *Theoretical Issues in Ergonomics Science*, *3*, 1–25.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 107–143.

Kirschenbaum, S. S. (1992). Influence of experience on information-gathering strategies. *Journal of Applied Psychology*, *77*, 343–352.

Kirschenbaum, S. S., & Arruda, J. (1994). The effects of graphic and verbal probability information on command decision making. *Human Factors*, *36*, 406–418.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 47–60.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*, 65–99.

Mahan, R. P., Kirschenbaum, S. S., Jilg, J. M., & Marino, C. (1998). Supporting diagnostic judgment during delayed-feedback decision making through reliability feedforward techniques. *Journal of Cognitive Ergonomics*, *2*, 217–242.

Nadav-Greenberg, L., & Joslyn, S. L. (2009). Uncertainty forecasts improve decision making among non-experts. *Journal of Cognitive Engineering and Decision Making*, *3*, 209–227.

Nadav-Greenberg, L., Joslyn, S. L., & Taing, M. U. (2008). The effects of uncertainty visualizations on decision making in weather forecasting. *Journal of Cognitive Engineering and Decision Making*, *2*, 24–47.

Pang, A. T., Wittenbrink, C. M., & Lodha, S. K. (1997). Approaches to uncertainty visualization. *The Visual Computer*, *13*, 370–389.

Slocum, T., Cliburn, D., Feddema, J., & Miller, J. (2003). Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance. *Cartography and Geographic Information Science*, *30*, 299–317.

Urick, R. J. (1983). *Principles of underwater sound* (3rd ed.). New York, NY: McGraw-Hill.

William, L. T., Scott, D. S., Marie, S. B., & Stephen, M. K. (2009). Two forms of spatial imagery: Neuroimaging evidence. *Psychological Science*, *20*, 1245–1253.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, *21*, 179–217.

Susan S. Kirschenbaum is an engineering psychologist in the Combat Systems Department of the Naval Undersea Warfare Center Division, Newport, Rhode

Island, and the Human Systems Integration (HSI) Lead for NUWCDIVNPT. Dr. Kirschenbaum received her PhD in experimental psychology from the University of Rhode Island in 1985. She has continuing interests in expertise and in the effects of uncertainty and other information variables on decision making. Applications for her research are in the design of systems for decision support, human-computer interaction, and display design. She was the Conference Chair for the 2004 and 2006 Undersea HIS Symposia. She is a Fellow of the American Psychological Association, was the 1996 recipient of the NUWC Independent Research Award, and has been a Visiting Scientist in both the United States and Australia.

J. Gregory Trafton leads a research team at the National Center for Applied Research in Artificial Intelligence of the U.S. Naval Research Lab in Washington, DC. Dr. Trafton is also a research professor at George Mason University in Fairfax, VA. Dr. Trafton earned his PhD in psychology from Princeton University. Current research interests include cognitive robotics and human-robot interactions; errors, interruptions, and resumptions; the cognition of complex visualizations; graph comprehension; and spatial cognition. He is the author of numerous scholarly publications.

Christian D. Schunn is an associate professor of psychology at the University of Pittsburgh, with secondary appointments in intelligent systems and learning sciences and policy. He received his PhD in cognitive psychology in 1995 from Carnegie Mellon University. His research ranges over spatial problem solving, innovative cognition, neuroscience of learning, peer instruction, science education, and design-based learning.

Susan B. Trickett is a cognitive psychologist in Denver, Colorado. She received her PhD from George Mason University.

Date received: January 28, 2010
Date accepted: June 17, 2013