Defence Research and
Development Canada

Recherche et développement
pour la défense Canada

DEFENCE **R&D** DÉFENSE

Representing data quality on naval tactical displays

Sharon McFadden
Defence R&D Canada – Toronto

Annie Li
University of Waterloo

Kevin Trinh
Defence R&D Canada – Toronto

Canada

Representing data quality on naval tactical displays

Sharon McFadden
Defence R&D Canada - Toronto

Annie Li
University of Waterloo

Kevin Trinh
Defence R&D Canada - Toronto

Defence R&D Canada – Toronto

Principal Author

Original signed by Sharon McFadden

....

Sharon McFadden

Defence Scientist

Approved by

Original signed by Justin Hollands

....

Justin Hollands

Head, Human Systems Integration Section

Approved for release by

Original signed by K.C. Wulterkens

....

K.C. Wulterkens

for Chair, Document Review and Library Committee

Abstract

Multi-Source Data Fusion (MSDF), as developed under COMDAT, provides an assessment of the reliability of the estimate of the fused track's attributes and position. Part of the human factors work under COMDAT has been to investigate methods of representing the quality of the MSDF-generated tracks to the operator. The research reported in this paper is concerned with the potential impact of different representations of data quality or uncertainty on the visibility of tactical symbols and the intuitiveness of the different representations. Three methods of representing data quality were investigated: a variably filled bar presented beside the tactical symbols, different diameter rings that encompassed the tactical symbols, and varying the saturation of the tactical symbols. In Experiment 1, a visual search task was used to compare the accuracy and speed with which participants could locate multiple instances of each of the tactical symbols without any representation of data quality and when each of the three methods tested were added. Experiment 2 examined the ability of participants to quickly and consistently interpret the quality of the data, represented by the tactical symbol, using the three different methods. The results indicated that the bar interfered least with people's ability to locate the tactical symbols, but the saturation method was most consistently interpreted. A second set of experiments looked at applying the saturation coding to the bars and rings instead of the tactical symbol. Redundant saturation coding of the bars and rings had no effect on tactical symbol visibility and did not improve the consistency of interpretation of data quality. The basic recommendation was that a small independent symbol, such as the bar, was the preferred method for representing data quality. However, operators should have the option of turning it off if the display appears too cluttered. If a different symbol shape is chosen, its intuitiveness should be assessed prior to implementation. Further research is required to improve our understanding of the number of levels of data quality operators can use effectively.

Résumé

La fusion de données de multiples sources (*Multi-Source Data Fusion*, MSDF), telle que mise au point dans le cadre du COMDAT, fournit une évaluation de la fiabilité de l'estimation des attributs et de la position des trajectoires. Les travaux sur les facteurs humains menés dans le cadre du COMDAT ont en partie consisté en l'étude de méthodes de représentation pour l'opérateur de la qualité des trajectoires générées par la MSDF. Les travaux de recherche dont il est fait état dans le présent rapport abordent l'incidence potentielle de différentes représentations de la qualité des données ou l'incertitude rattachée à la visibilité des symboles tactiques et le caractère intuitif de différentes représentations. Trois méthodes de représentation de la qualité des données ont été étudiées : une barre plus ou moins remplie placée à côté des symboles tactiques, des anneaux de différents diamètres autour des symboles tactiques et la variation de la saturation des symboles tactiques. Dans la première expérience, une tâche de recherche visuelle a été utilisée pour comparer l'exactitude et la vitesse avec lesquelles les participants pouvaient localiser de multiples occurrences de chacun des symboles tactiques sans représentation de la qualité des données et dans le cas de chacune des trois méthodes éprouvées. Dans la deuxième expérience, on a examiné l'aptitude des participants à interpréter rapidement et de manière uniforme la qualité des données, représentées par le symbole tactique, pour chacune des trois méthodes différentes. Les résultats indiquent que c'est la barre qui nuit le moins à l'aptitude des gens à localiser les symboles tactiques, mais la méthode de la saturation est celle qui permet l'interprétation la plus uniforme. Dans un deuxième ensemble d'expériences, on a examiné l'application d'un codage en saturation des barres et des anneaux plutôt que des symboles tactiques. Le codage redondant en saturation des barres et des anneaux n'a aucun effet sur la visibilité des symboles tactiques et n'améliore aucunement l'uniformité de l'interprétation de la qualité des données. La recommandation fondamentale est de préférer un petit symbole indépendant comme la barre pour la représentation de la qualité des données. Cependant, les opérateurs devraient avoir le choix de le supprimer de l'affichage si celui-ci devenait trop encombré. Si un symbole de forme différente était retenu, il faudrait en évaluer le caractère intuitif avant la mise en œuvre. D'autres recherches sont nécessaires pour améliorer notre compréhension du nombre de niveaux de qualité des données que les opérateurs peuvent efficacement utiliser.

Executive summary

Representing data quality on naval tactical displays

S. McFadden; A. Li; K. Trinh; DRDC Toronto TR 2007-032; Defence R&D Canada – Toronto; January 2008.

Introduction or background: The research reported in this paper was conducted under the COMmand Decision Aiding Technology (COMDAT) (11bg) Technology Demonstrator Project (TDP) in support of the Halifax Class Modernization Command and Control System (HMCCS) Programme. Multi-Source Data Fusion (MSDF), as developed under COMDAT, provides an assessment of the reliability of the estimate of the fused track's attributes and position. Part of the human factors work under COMDAT has been to investigate methods of representing the quality of the MSDF-generated tracks to the operator on the tactical display. The research reported in this paper is concerned with the potential impact of different representations of data quality or uncertainty on the visibility of tactical symbols and the intuitiveness of the different representations. Three methods of representing data quality were investigated. The first method involved placing a second, smaller symbol, a bar, to the right of the tactical symbol. The amount of fill in the secondary symbol provided information about the quality of information represented by the tactical symbol. The second method involved annotating the symbol by adding a ring around it. The diameter of the ring provided information about the quality of information represented by the tactical symbol. The third method involved modifying the tactical symbol by changing its saturation as a function of the quality of information. Four experiments were carried out. Experiment 1 measured the impact of the different methods for representing data quality on the visibility of tactical symbols using a visual search task. Experiment 2 assessed the intuitiveness of the different methods for representing data quality. Experiments 3 and 4 were similar to the first two experiments except they examined the effects of using redundant saturation coding of the bar and rings instead of the tactical symbol.

Results: Relative to the control condition, the bar interfered least with search performance and the saturation method interfered most. On the other hand, estimate of data quality were most consistent with the saturation method. Thus, no one method was optimal across the two tasks. In experiments three and four, there was no significant advantage to using redundant saturation coding.

Significance: The basic recommendation was that a small independent symbol such as the bar was the preferred method for representing data quality. The bar interfered least with searching for tactical symbols and most participants interpreted it in the same way. Methods that modify the tactical symbols, such as saturation coding, should be avoided. If a symbol other than the bar symbol is being considered, its intuitiveness should be assessed prior to implementation. The chosen symbol should not be larger than the bar and the different forms of the symbol should be clearly discriminable. Finally, operators should have the option of removing the data quality symbol from the display to reduce clutter.

Future plans: Further research is required to improve our understanding of the number of data quality levels operators can use effectively.

Sommaire

Representing data quality on naval tactical displays

S. McFadden; A. Li; K. Trinh; DRDC Toronto TR 2007-032; R & D pour la défense Canada – Toronto; Janvier 2008.

Introduction ou contexte : Les recherches ici présentées ont été menées dans le cadre du Projet de Démonstrateur de la Technologie (PDT) d'Aide aux Décisions de Commandement (COMDAT) (11bg) à l'appui du Programme de modernisation du système de commandement et de contrôle pour la classe Halifax (*Halifax Class Modernization Command and Control System Programme*, HMCCS). La fusion de données de multiples sources (*Multi-Source Data Fusion*, MSDF), telle que mise au point dans le cadre du COMDAT, fournit une évaluation de la fiabilité de l'estimation des attributs et de la position des trajectoires. Les travaux sur les facteurs humains menés dans le cadre du COMDAT ont en partie consisté en l'étude de méthodes de représentation pour l'opérateur de la qualité des trajectoires générées par la MSDF dans l'affichage tactique. Les travaux de recherche dont il est fait état dans le présent rapport abordent l'incidence potentielle de différentes représentations de la qualité des données ou l'incertitude rattachée à la visibilité des symboles tactiques et le caractère intuitif de différentes représentations. Trois méthodes de représentation de la qualité des données ont été étudiées. La première méthode consiste à placer un deuxième symbole plus petit, une barre, à la droite du symbole tactique. Plus ou moins rempli, ce symbole secondaire fournit l'information sur la qualité de l'information représentée par le symbole tactique. La deuxième méthode consiste à annoter le symbole en ajoutant un anneau autour. Le diamètre de l'anneau fournit l'information sur la qualité de l'information représentée par le symbole tactique. La troisième méthode consiste à modifier le symbole tactique en changeant sa saturation en fonction de la qualité de l'information. Quatre expériences ont été effectuées. Dans la première expérience, on a mesuré l'incidence des différentes méthodes de représentation de la qualité des données sur la visibilité des symboles tactiques en utilisant une tâche de recherche visuelle. Dans la deuxième expérience, on a évalué le caractère intuitif des différentes méthodes de représentation de la qualité des données. Les troisième et quatrième expériences étaient similaires aux deux premières sauf qu'elles permettaient d'examiner les effets de l'utilisation d'un codage redondant en saturation des barres et des anneaux plutôt que du symbole tactique.

Résultats : Par rapport à la situation témoin, c'est la barre qui nuit le moins au rendement de la recherche et la méthode de la saturation qui nuit le plus. D'autre part, l'estimation de la qualité des données est la plus uniforme par la méthode de la saturation. Ainsi, aucune méthode individuelle n'offre un rendement optimal pour les deux tâches. Dans les expériences trois et quatre, aucun avantage n'a été constaté à l'utilisation d'un codage redondant en saturation.

Portée : La recommandation fondamentale est de préférer un petit symbole indépendant comme la barre pour la représentation de la qualité des données. C'est la barre qui nuit le moins à la recherche des symboles tactiques et la plupart des participants l'ont interprétée de la même manière. Les méthodes modifiant les symboles tactiques, comme le codage en saturation, devraient être évitées. Si un symbole de forme différente était retenu, il faudrait en évaluer le caractère intuitif avant la mise en œuvre. Le symbole retenu ne devrait pas être plus grand que la barre et les différentes formes du symbole devraient être nettement distinguables. Enfin, les

opérateurs devraient avoir le choix de supprimer de l'affichage le symbole de qualité des données pour réduire l'encombrement de l'affichage.

Recherches futures : D'autres recherches sont nécessaires pour améliorer notre compréhension du nombre de niveaux de qualité des données que les opérateurs peuvent efficacement utiliser.

This page intentionally left blank.

Table of contents

List of figures

List of tables

Acknowledgements

This page intentionally left blank.

Introduction

Background

The purpose of the COMmand Decision Aiding Technology (COMDAT) Technology Demonstrator Project (TDP) is to research and demonstrate Multi-Source Data Fusion (MSDF) technologies and carry out human factors studies in support of the Halifax Class Modernization Command and Control System (HMCCS) Programme in the areas of battle space awareness. One of the goals of the human factors studies is to develop improved guidelines for designing the Operator Machine Interface (OMI) for naval command and control systems.

Multi-Source Data Fusion (MSDF), as developed under COMDAT, fuses kinematic (position) and attribute (platform, affiliation, etc.) data from one or more sources into a single track. Part of the output of the fusion process is an assessment of the reliability of the fused track's attributes and position. Knowledge of these estimates of reliability could assist the operator in deciding what action to take in regards to the track. Thus part of the human factors work under COMDAT has been to investigate ways of representing the quality of the MSDF-generated tracks to the operator. Moreover, recent research (Waldron, Duggan, Patrick, Banbury, and Howes 2005) suggests that failure to provide this type of information can impact task performance and ability to maintain situation awareness.

One possible method is to present the information in the form of a table or graph when the operator selects or hooks a track. While this method provides the operator with the detail necessary to estimate the quality of a particular track, it does not give the operator a quick appreciation of the relative quality of the information for all of the tracks, as represented by the tactical symbols (Kirschenbaum and Arruda 1994). Annotating or modifying the track symbology integrates the quality of the information with the symbol representing that information and can potentially provide the operator with such an overall appreciation. Thus, the COMDAT project identified the requirement for data quality or uncertainty symbology to help inform the operator about track quality.

Representation of uncertainty

In recent years, there has been a growing interest in uncertainty in decision making in terms of how to reduce it (Hutton 2004) and how to represent it (Harrower 2003, Howes and de Bruijn 2005). The current work is concerned with the representation of uncertainty or the requirement to express the inherent uncertainty associated with information being presented. The intent is to convey to the user that the information as displayed does not necessarily represent the truth. There is some variability associated with it due to the incomplete, imperfect or out-or-date data on which it is based. Thus, the intent is really to increase uncertainty or more correctly to try to instil an appropriate level of uncertainty about the accuracy of the perceived information.

Methods for displaying this type of uncertainty have largely been examined by the measurement community and more recently with the advent of Geographic Information Systems (GIS), by the geographic community. According to Harrower (2003), the GIS community has "made great advances in defining, measuring, modeling, and visualizing uncertainty". Each of these steps is

important to the ultimate usefulness of the representation of uncertainty or data quality to the end user. As is often the case, there has been less effort in determining the value of different methods of representing uncertainty although progress is being made in that area.

Since the goals of users of GIS information are often different than users of tactical displays, many of the findings in this area may not be directly relevant. Probably the most relevant are the studies on different methods for incorporating uncertainty representations. Evans (1997) compared four methods of depicting data quality – static separate displays, a static integrated display, animated non-controllable "flicker maps", and interactive toggle displays. The study compared the effectiveness of the different methods for displaying a land use map and a "95% reliable map" that showed only the land use of areas that could be classified with 95% accuracy. With the flicker map, the display alternated between the two maps at a rate of 4 frames per second. The results showed that participants preferred and performed best with the static integrated display and the flicker display. Edwards and Nelson (2001) also found that integrated displays worked better than separated displays and that traditional verbal statements worked least well. Kirschenbaum and Arruda (1994), in a study on decision making in a sonar target tracking task, also found that an integrated display with a spatial representation of uncertainty was superior.

Overall, the research supports the decision to provide a representation of data quality on the tactical display rather than requiring the operator to access the information by calling up a separate display or by querying a specific track. There are many different ways that data quality information can be integrated into the primary display. Howes and de Bruijn (2005) provide a summary of many of the different approaches that have been tried and what is known about their relative merits. Unfortunately, as indicated by Harrower (2003), many of the studies do not include objective or even subjective evaluations of different methods.

An exception is a study by Finger and Bisantz (2002) which compared different visual representations of uncertainty. In that study, the authors investigated the utility of using variably blurred icons to convey different probabilities of uncertainty about the friendliness or hostility of an object. Pairs of icons that could represent the identity of objects as hostile and friendly were blurred and blended to varying degrees to form series of thirteen icons. Each series represented a range of probabilities from 0 to 100%. Participants ordered each set of images from least (0%) to most (100%) friendly or from least to most hostile. They then rated the friendliness (or hostility) of each icon in the series along a continuous scale. On average, people could consistently and accurately order the icons although the order was sometimes in the opposite direction (the nominally hostile icon was seen as friendly). The set of blurred icons were then used in a task where participants had to identify objects as friendly or hostile based on probabilistic information represented by the blurred icons alone, the blurred icons plus a numerical probability, or just the non-blurred icons plus a numerical probability. The results showed that the participants tended to be less conservative and faster in the icon only condition which led to the identification of more objects, but lower accuracy compared to when the numerical information was available. The results also indicated that people did not make use of the available range of probabilities provided, but grouped them into one of three categories, high probability, moderate probability, and low probability that the target was friendly (or hostile).

Although the members of each series were discriminable and could be ranked, the study did not determine if people were able to uniquely associate a specific level with a specific icon. Could

people make use of that many different probability levels? Certainly, the tendency for people in the icon only group to be less conservative could have been due to an inability to accurately identify the intermediate levels of blur. Guidelines for display design usually recommend using no more than two or three levels with coding dimensions such as brightness and size where discrete names cannot be applied to the different levels (Unger Campbell 2004). Blur would certainly fall in this category. With other dimensions such as hue or slope, the limit is around 6 to 8 levels.

Some of the results of the previous study (Finger and Bisantz 2002), are consistent with a small but growing literature that shows the importance of the form in which the information is presented on decision making. In it, inter-participant consistency in ranking the symbols was significantly higher when people were asked to rank the symbols as more or less friendly rather than more or less hostile. More direct support for the importance of context on decision making is shown in a study by Banbury, Selcon, Endsley, Gordon, and Tatlock (1998). They assessed a pilot's willingness to shoot a target based on the form of the estimates of target identification reliability provided by a decision support system. The estimate was framed either as degree of confidence (5 levels between 61-97%) or degree of uncertainty (5 levels between 3-39%). A second parameter was the presence and type (hostile or friendly) of alternative hypothesis. Except for the highest and lowest levels of uncertainty or confidence, the number of shots taken declined when the alternative hypothesis was friendly as opposed to hostile or not present. Also, response time slowed significantly when the confidence level was very high (uncertainty very low). Fastest response times occurred when there was no alternative hypothesis. The data also suggested that people tended to categorize the different estimates into two groups acceptable (above 91% or below 9%) and unacceptable. There was no overall effect of framing although there was the suggestion that participants were translating the uncertainty estimates into confidence estimates.

Current study

In the operations room of the frigate, information on the tactical display must be processed quickly and accurately. Thus, it is important that the selected representation of data quality has minimal negative impact on tactical picture processing and be consistently and quickly interpreted. To date, there appears to have been little or no direct examination of the impact of the form of visual representation on increased clutter and consistency of interpretation. The current research under COMDAT is an initial effort to address this deficiency. Prior to the experiment reported in this paper, several different designs for representing data quality of MSDF tracks on a tactical display were developed (Lockheed Martin Canada Inc. 2001). Four of these (Table 1) were selected for further evaluation of the intuitiveness of their representation of data quality (uncertain versus certain) and their representation of types of data quality (e.g., position, time, affiliation etc.) (Unger Campbell and Baker 2003). The study rated the preference of Navy Subject Matter Experts (SMEs) for the four symbol shapes and asked them what kind of certainty and what level of certainty they thought each symbol represented. SMEs preferred the dot representation (1d), but the slider (1a) elicited the most consistent response in terms of level of certainty represented.

Table 1: Symbol sets used in earlier certainty evaluation. In each set, the first three symbols nominally represented most, mid-level and least certain. The four bars(c) were intended to represent different types or sources of certainty. Thus they could differ in height. The 'no dot' condition (d) could represent least or most certain.

a) Slider (Also known as draining bucket)

b) Sights (Also known as gun sights)

c) Bars

d) Dots
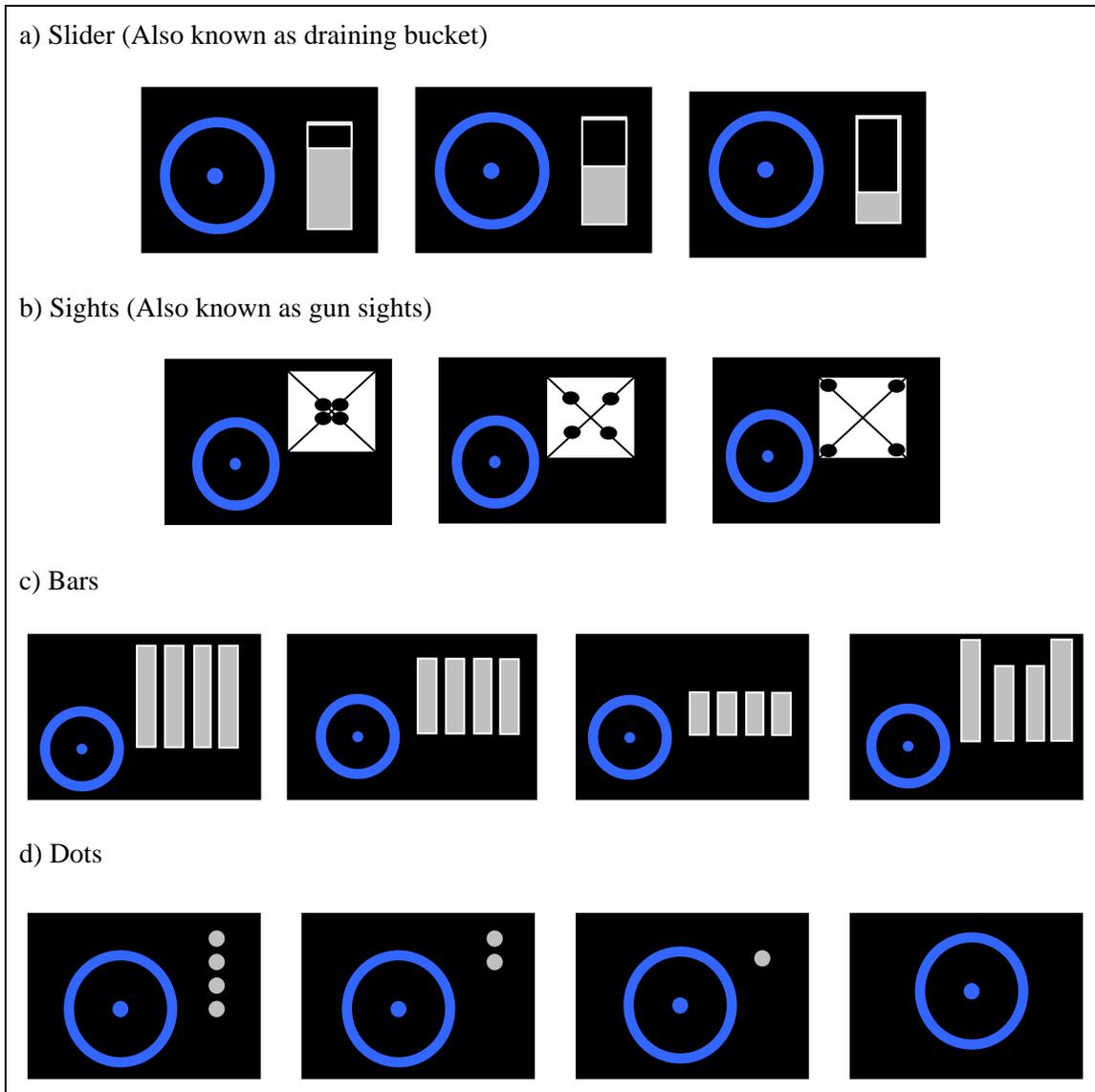
This evaluation primarily examined the subjective impression of operators. It did not look at how quickly or accurately an operator could assess the quality of a track in the operational context using the symbology. Also, it did not address the impact of the additional clutter associated with using a symbol to represent data quality. The increased clutter may make it more difficult for the operator to detect new tracks or to locate specific tracks. Also, it may disrupt the ability to assess the overall tactical picture. The alternative is to modify the tactical symbol in some way.

The experiments reported here attempted to address both of these issues. Initially, two experiments were carried out. Experiment 1 assessed the impact of three different data quality method on quickly and accurately locating specific data quality symbols. Experiment 2 examined the ability of participants to quickly and consistently interpret the quality of the data represented by the tactical symbol using the same three methods. The three different methods were a variably filled bar (a slight variant of the slider in the previous study), different diameter rings, that encompassed the tactical symbols, and saturation coding of the tactical symbols. As discussed below, the three methods were chosen because they were perceived as having a differential impact on display clutter and symbol visibility and the possibility of being interpreted consistently. In keeping with human factors guidelines discussed earlier, only three different levels of data quality were represented with each method.

In the bar condition, a small bar appeared to the right of the symbol. Data quality was indicated by the amount the bar was filled. One interpretation would be that the more filled the bar is, the better the quality of data represented by the tactical symbol. The bar is representative of methods using an independent symbol. An independent symbol should have minimal impact on the appearance of the tactical symbol, but the number of symbols on the display doubles making the display more cluttered. Other possibilities are shown in Table 1. The bar was chosen because it had reasonable operator acceptance and was most consistently interpreted in the Unger Campbell and Baker (2003) study.

In the ring condition, data quality was represented by the diameter of a ring that surrounded the tactical symbols. One interpretation would be that the smaller the ring the better the data quality. The ring is representative of a method that annotates the tactical symbol. The symbol itself is unchanged, but the annotation may reduce the similarity of multiple examples of the same symbol and increase the similarity of different types of symbols. Since the ring appears as part of the symbol, there should not be the same increase in clutter as with an independent symbol. Another example would be the standard error bars used in graphical displays. In that case, the shorter the line is, the more certain the estimate of the mean. The ring was chosen because operators already use rings to represent position uncertainty of tracks on tactical displays. However, those rings encompass the actual area on the tactical display where a track might be located.

In the saturation condition, the saturation (colourfulness) of the symbol colour indicated data quality. One interpretation would be that the more saturated or colourful the symbol the better the data quality. Saturation coding is representative of methods that modify the characteristics of the actual symbol. Such methods reduce the similarity of multiple examples of the same symbol along a specific dimension. However, they should not increase display clutter. The blurring used by Finger and Bisantz (2002) would be another example of this type of method.

The participant's task in Experiment 1 was to count the number of instances of a specific tactical symbol shape on the display. This counting task provides an estimate of how quickly and accurately the participant can extract information about the location and existence of contacts or tracks on a display (e.g. the number of hostile aircraft). The symbols were presented by themselves or they were annotated with one of the three different methods discussed above quality. In Experiment 2, participants were required to search for a specified tactical symbol and then estimate the quality of the underlying data using the representation of data quality associated with the specified symbol. Although only three levels of data quality were represented, participants were told to use a 5 point scale. It was felt that they would feel less constrained in

terms of defining the level of perceived data quality. It is important that when different operators look at a particular symbol, they all interpret the different forms of the data quality representation consistently and similarly. Moreover, a three point scale would not allow participants to take account of interactions between the tactical symbol shape and the data quality representation method. Thus, the perceived diameter of the circle might change when it surrounded a surface symbol as compared to an air symbol. Also, the saturation levels might change as a function of a symbol's location on the display.

The symbol set currently used by the Canadian Navy is the Naval Tactical Data System (NTDS). However, consideration is being given to introducing the MIL-STD-2525B (2525B) symbol set (Department of Defence 1999). The 2525B symbol set is much richer than the NTDS set and allows for more extensive annotation of the symbols. Thus the experiments looked at the impact of the different methods on both sets. In its basic form, each symbol represents the platform (air, surface, subsurface) and affiliation (friendly, hostile, unknown, and neutral (2525B only)) of the underlying target.

Experiments 1 and 2

Method

Participants

Eight participants, 3 males and 5 females, took part in the two experiments. They ranged in age from 18 to 50 with a mean age of 30. All participants had normal or corrected-to-normal vision based on self-report and normal colour vision as assessed by the Ishihara pseudoisochromatic plates. All participants signed an informed consent form approved by the Defence Research and Development Canada Human Research Ethics Committee (HREC) before participating in the experiments (DRDC HREC Protocol L508).

Conditions

Experiment 1 was a 2 (symbol set: NTDS and 2525B) by 4 (method for representing data quality: bar, saturation, rings, and control) within participant design. The design for Experiment 2 was identical except that there was no control condition. Participants carried out 72 runs (2 symbol sets by 4 methods by 9 symbol types) in Experiment 1 and 54 runs (2 symbol sets by 3 methods by 9 symbol types) in Experiment 2. The order in which the four methods were carried out was randomized across participants and symbol sets. Participants always completed Experiment 1 prior to Experiment 2.

Apparatus

Stimuli were presented on a 19" (0.48 m) diagonal ViewSonic Professional Series P817 monitor set to 1280 by 1024 pixels. All responses were collected via the numeric keypad on the computer keyboard. The chromaticity and luminance of each of the stimuli were measured with a Minolta CS100 chroma meter. The illuminance of the screen and keyboard were measured using a Hagner photometer in illuminance mode. Measurements were repeated at weekly intervals. The variability over time was less than .005 in x and y for the chromaticity coordinates and less than 10% for the luminances and illuminances.

Stimuli

The stimuli were 9 symbols from the NTDS set and the equivalent symbols from 2525B (Table 2). In both sets, geometric shape (circle, club, diamond, and square) is used to code affiliation and whether the shape is closed (surface), cut-off at the top (underwater), or at the bottom (air) determines platform. As well, affiliation is redundantly colour coded. The symbol sets differ primarily in that NTDS uses outlines and 2525B uses filled shapes. The visual angle of each symbol type is also shown in Table 2. The relative size of the symbols was based on the recommendations in MIL-STD-2525B for that symbol set and observation of operational displays for the NTDS symbol set.

Table 2: Colour-coded 2525B and NTDS symbol sets used in experiments. The visual angle of each symbol on the display in minutes of arc at a 60 cm viewing distance is shown below the symbol.

Symbol Set	Platform	Affiliation		
		Friendly	Hostile	Unknown
2525B	Air	40 by 40	40 by 40	40 by 50
	Surface	40 by 40	50 by 50	50 by 50
	Subsurface	40 by 40	40 by 40	40 by 50
NTDS	Air	20 by 46	23 by 46	20 by 40
	Surface	46 by 46	50 by 50	40 by 40
	Subsurface	20 by 46	23 by 46	20 by 40
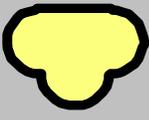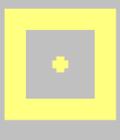
Except for the control condition, each symbol was annotated (bar or ring) or modified (saturation) (Table 3). In the bar condition, a small bar appeared to the right and slightly above the symbol. Data quality was indicated by the percentage that the bar was filled in. In the ring condition, the symbol was surrounded by a ring. Data quality was indicated by the diameter of the ring (see Table 3 for the visual angle of the bars and rings). In the saturation condition, the saturation (colourfulness) of the symbol colour was set to one of three levels. The method for determining

the different saturation levels is described in Annex A and the average chromaticity coordinates and luminance levels of the three different saturation levels are given in Table 4. The 1931 CIE chromaticity coordinates and luminance of the bar and ring were x = 0.281, y = 0.314, L = 71 Cd/m^2. The luminance of the display background was approximately 7 Cd/m^2.

Table 3: An example of each level of the three different data quality representations when associated with the 2525B surface friendly symbol. The visual angle of the bar and rings in minutes of arc at 60 cm viewing distance is shown below the symbol.
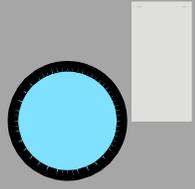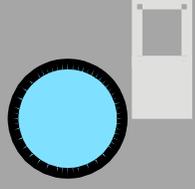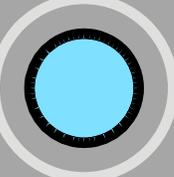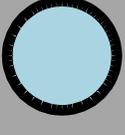
Data quality category	Method		
	Bar	**Rings**	**Saturation**
A	40 by 20	63	
B		69	
C		74	
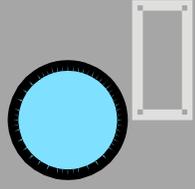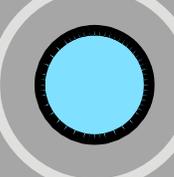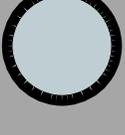
Table 4: CIE 1931 Chromaticity coordinates (x,y) and average luminance (Cd/m2) of each of the symbols at each of the saturation levels. The symbols in the other conditions have the same values as level A in the saturation condition.

Data quality category	Affiliation		
	Friendly	Hostile	Unknown
A	0.211, 0.262, 64	0.411, 0.326, 40	0.358, 0.442, 92
B	0.241, 0.283, 62	0.325, 0.313, 39	0.315, 0.373, 78
C	0.263, 0.299, 61	0.291, 0.308, 40	0.291, 0.331, 70

Display

On each trial participants were shown a display containing 50 symbols randomly placed in the cells of an imaginary 8 (vertical) by 8 (horizontal) array and presented against a dark grey background. The symbols were randomly offset between 0 and 90 minutes of arc in both x and y to make the display appear unordered and it usually resulted in a few of the symbols overlapping. The nominal size of the grid was 23.5 by 23.5 degrees of arc at a viewing distance of 60 cm. In Experiment 1, the 50 symbols were made up of between 3 and 6 target symbols and 44 and 47 distractor symbols. In Experiment 2, there was only 1 target symbol and 49 distractor symbols.

Tasks

The participants' task in Experiment 1 was to count the number of target symbols on the display as quickly and accurately as possible and to indicate their response by pressing the appropriate number key. In Experiment 2, the participants were required to search for one instance of the target symbol and then, based on their interpretation of the associated bar, ring or saturation level, indicate the quality of the information represented by that symbol on a scale of 1 to 5 by pressing the corresponding number on the keyboard.

Procedure

The experiments took approximately 5 sessions of about 90 minutes each (including breaks). In session one, the participant read the general information sheet and signed the informed consent form. Any questions regarding the general design of the experiment were answered. Next the participant's colour vision was tested. If the participant passed the colour vision test, they carried out the practice session followed by the first two conditions for Experiment 1. In the next two sessions, the participants completed Experiment 1 and in the last two sessions they carried out Experiment 2.

All experimental sessions were carried out under relatively low ambient illumination to maximize the visibility of the symbols. The light falling on the screen was approximately 4 Lux and the

light falling on the keyboard 12 Lux. Participants were seated in an adjustable chair in front of an adjustable keyboard at a distance of approximately 60 cm from the computer screen.

The practice session was designed to give them familiarity with the symbol set and the counting task. In it, participants were given four trials with each of the 2525B and NTDS symbols as targets. The number of targets differed on each of the four trials.

Each condition was composed of 9 runs. At the start of each run, a screen specifying the target and distractor types for that run was presented. In all cases only the basic symbols, but not the data quality information, was shown on the introduction screen. This display remained on until the participant signalled with a keystroke to continue. Next, 4 practice trials were presented (one with each of the different numbers of targets) with feedback in the form of a plus for a correct response or a minus for an incorrect response, followed by 16 experimental trials (again with trial-by-trial feedback). Displays remained on the screen until a response was made. The participants were instructed to count the number of targets on the screen and report the result using the numbers on the keyboard.

The procedure for a run was identical for Experiment 2 except that participants received three practice trial (one for each level of data quality) and 12 experimental trials per run. They were told that the bar (or saturation or ring) gave an indication of the quality of the data represented by the target symbol. When each test screen appeared, their task was to locate the specified target for that run and then, based on their interpretation of the additional information, rank the quality of the underlying data represented by that symbol on a scale of 1 to 5 where 5 meant good quality and 1 meant very poor quality. In other words, how confident were they that the information that was conveyed by the target symbol about the target's current position, platform (air, surface, subsurface), and affiliation (friendly, hostile, unknown) was up to date and/or accurate based on the appearance of the data quality cue. They were further instructed that there were no right or wrong answers, but that they should be consistent in their interpretation of the data quality cues. For example, if they saw the same saturation on different trials they should give it a similar rating and they should try to avoid rating a particular cue as 5 on one trial and 1 on a different trial.

The order of presentation of the different targets within a condition and the order of the conditions within a symbol set were randomized across participants. The two symbol sets were always presented in separate sessions.

Results

Experiment 1 – counting task

The results of interest were median response time per run for correct responses and percent errors per run. The median response time per run was used to control for outliers. Although response times were relatively normally distributed (usually skewness was less than 1.5 across conditions), using log response times did reduce skewness to less than 1 across all conditions. Thus median log response times were used in all the analyses. Prior to conducting the main analysis, the effect of number of targets on response time was analysed to determine whether response time per trial or response time per target was the most accurate measure of response time. The results indicated that response time did not increase as number of targets increased. Thus, median log response

time per trial was used. The initial analysis examined the effect of symbol set and method on accuracy and response time. Since the symbol shapes differed in each symbol set, the effect of method and symbol shape on accuracy and response time was analysed separately for each symbol set. Both analyses used a repeated measures analysis of variance. Main effects and interactions were included in the models and the Scheffe test was used for post-hoc analysis. A significance level of 0.01 was used in all cases. Since there is often a trade off between response time and accuracy in visual search tasks, multivariate analyses of log median response times and accuracy were also carried out. As can be seen from Table 5, there was a significant effect of method on accuracy and response time and overall. The post hoc analyses of method indicated that accuracy was significantly lower and response time significantly slower in the saturation condition relative to the remaining methods (Figure 1). As well, response time was significantly slower with the ring and bar methods relative to the control condition. There was no significant interaction between symbol set and method.

Table 5: Summary of repeated measures analyses of variance for effect of symbol set and method on accuracy and response time separately and a multivariate analysis of overall effect of both. The Wilks' Lambda statistic is reported for the multivariate analysis.

Source	Univariate / multivariate degrees of freedom	F values (P <0.01)		
		Accuracy	Response time	Multi-variate
Symbol set	1,7 / 2,6	n.s.*	n.s.	17.3
Method	3,21 / 6,40	24.6	60.1	20.4

*not significant
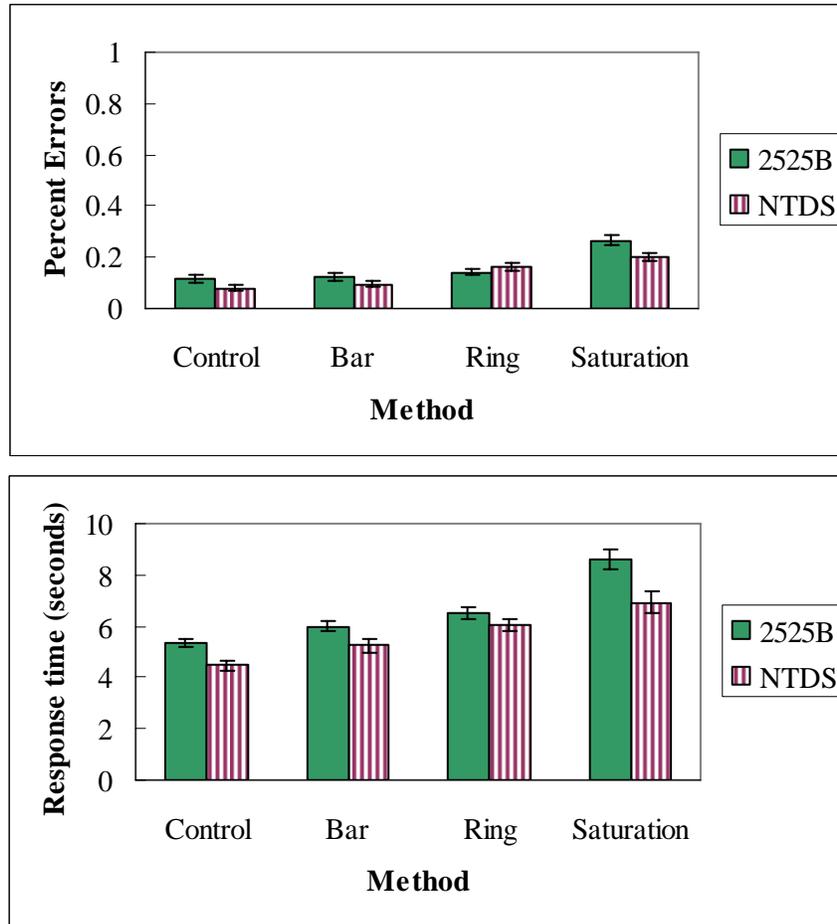
Figure 1: Effect of symbol set and method on percentage of errors and response time. The error bars indicate the standard error of the mean.

The separate analysis of each symbol set (Table 6) showed a similar pattern of results for method. It also showed a significant effect of symbol shape and a small but significant interaction between symbol shape and method with each symbol set. The pattern of results for symbol shape were similar to the results in another study (McFadden, Jeon, Li, and Minniti 2007) that focused on the effect of symbol type and symbol set on visual search performance. Performance with the surface symbols, particularly the surface friendly, was faster and more accurate. An examination of the interaction between symbol shape and method indicated that it was primarily due to the effect of the ring method on the surface friendly symbol. With most symbols, performance was poorest with the saturation method. However as indicated in Figure 2, performance tended to be less accurate and slower with the ring method with the surface friendly symbol, especially for the NTDS set.

Table 6: Summary of repeated measures analyses of variance for effect of symbol set and method on accuracy and response time separately and a multivariate analysis of overall effect of both. The Wilks' Lambda statistic is reported for the multivariate analysis.

Symbol set	Source	Univariate / multivariate degrees of freedom	F values (P <0.01)		
			Accuracy	Response time	Multi-variate
2525B	Method	3,21 / 6,40	22.1	50.4	16.8
	Symbol	8,56 / 16,110	8.2	43.8	14.9
	Method by symbol	24,168 / 48,334	3.0	4.7	3.6
NTDS	Method	3,21 / 6,40	11.6	31.5	11.9
	Symbol	8,56 / 16,110	7.5	177.8	34.8
	Method by symbol	24,168 / 48,334	2.1	7.5	4.3
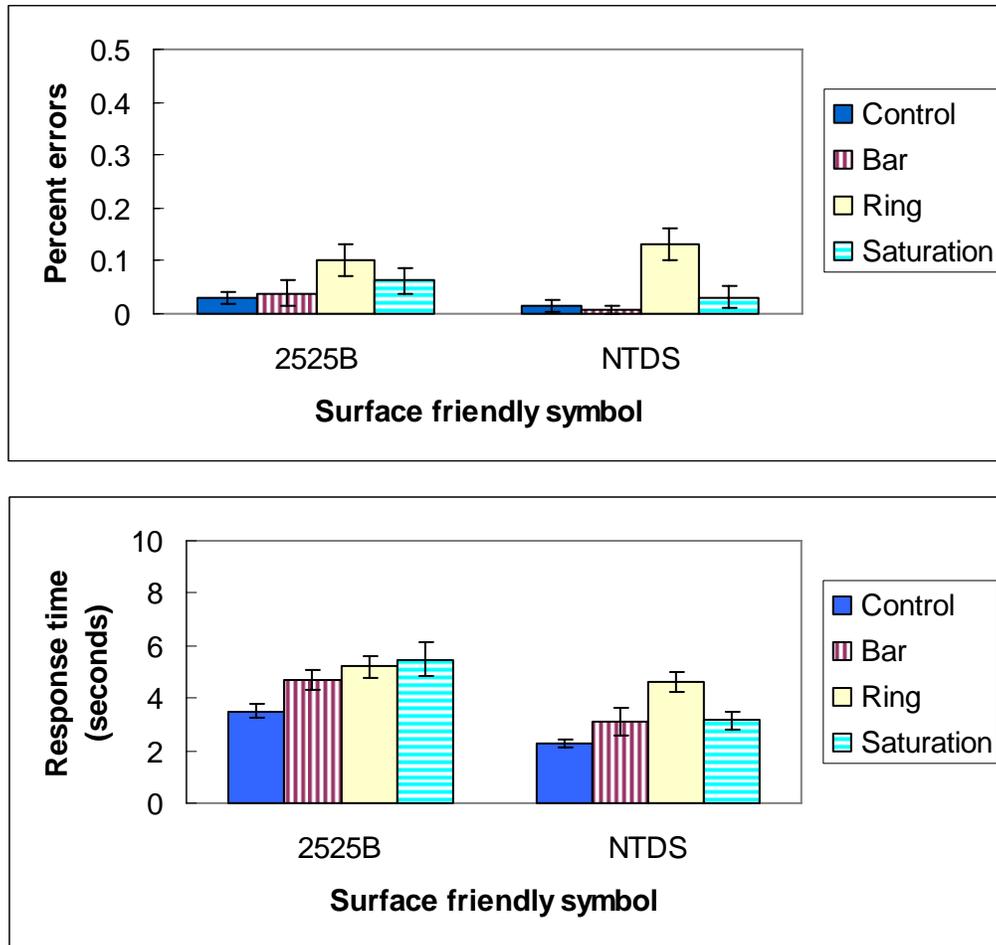
Figure 2: Effect of data quality representation and symbol set on accuracy and response time for the surface friendly symbol from both symbol sets.

Experiment 2 – rating task

The participant's responses in Experiment 2 were their estimations of the quality of the information represented by the symbol based on their interpretation of the data quality metric. Response times were also recorded. Initially, the median rating and median log response time for each combination of symbol set, method, symbol, and representation were calculated. These medians were used in all the remaining analysis.

Ideally, all participants would have interpreted the data quality representations in the same way and would have used the whole range of the scale assigning 1 or 2 to the best representation of low data quality, 2-4 to the best representation of intermediate data quality, and 4 or 5 to the best representations of high data quality. As Figure 3 shows, this was not the case. Only the saturation method appeared to elicit a consistent response across all participants.
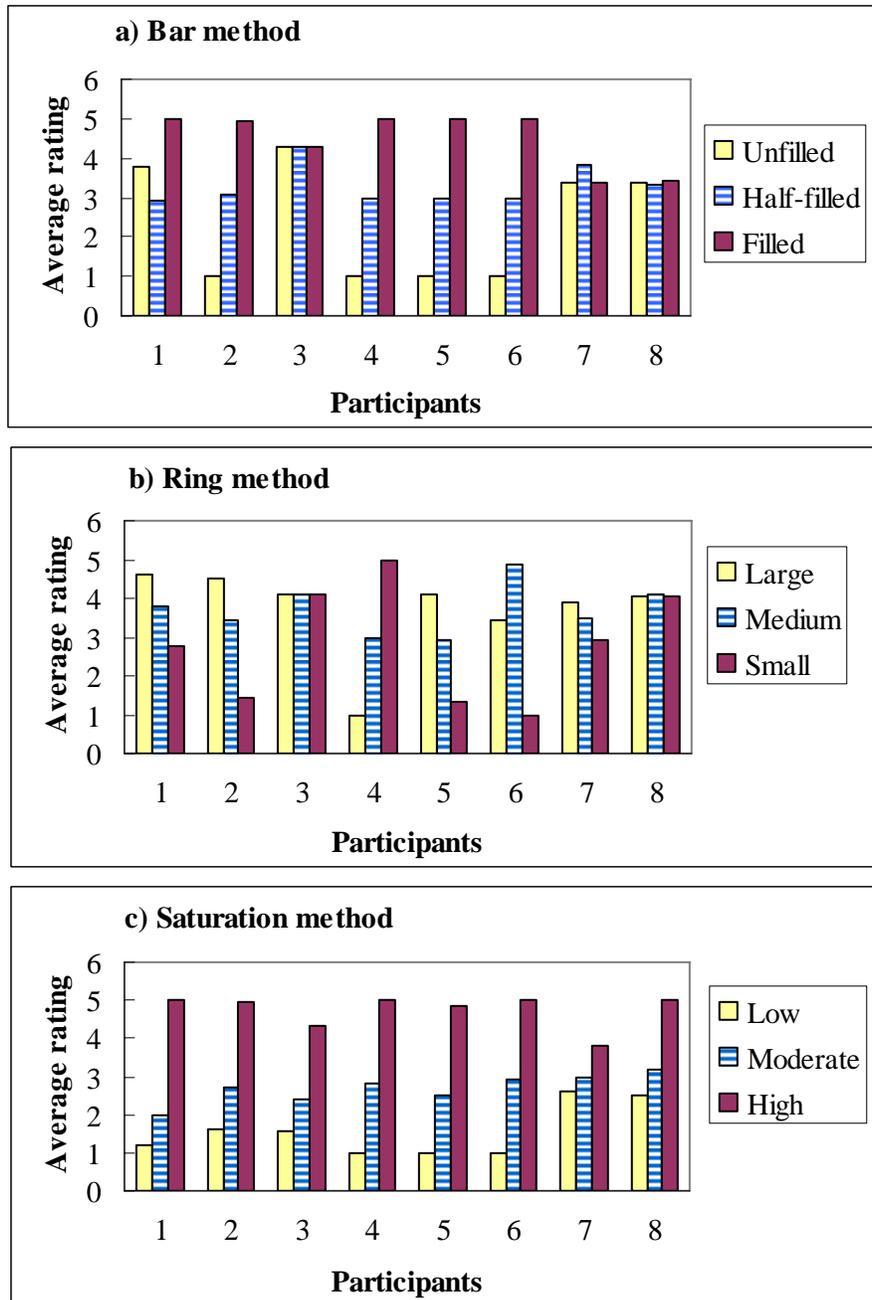
Figure 3: Average rating by each participant for the three forms of each of the three methods. The ratings were always between 1 and 5 and a higher rating indicated greater certainty.

Since the three representations were not necessarily equivalent across the different methods, a separate analysis of the effect of symbol set, and representation on rating response was carried out for each method. A similar analysis was carried out on the median log response times. In addition, the effect of symbol shape and representation on response time was examined for each of the

symbol sets. For all analyses, a repeated measures analysis of variance was used. Main effects and interactions were included in the models.

The ratings for each representation were significantly different for the saturation ($F(2,14) = 74.4$) and bar ($F(2,14) = 8.4$) methods, but not for the ring method (Figure 4). The post-hoc analysis showed that the rating for each representation was significantly different from the remaining two in the saturation method with the high saturation representation being associated with high certainty. In the bar method, only the ratings for the filled and unfilled bar were significantly different. The filled bar was associated with good data quality and the unfilled bar with moderate to low data quality. There was no effect of symbol set and no interaction between symbol set and representation for any of the methods.
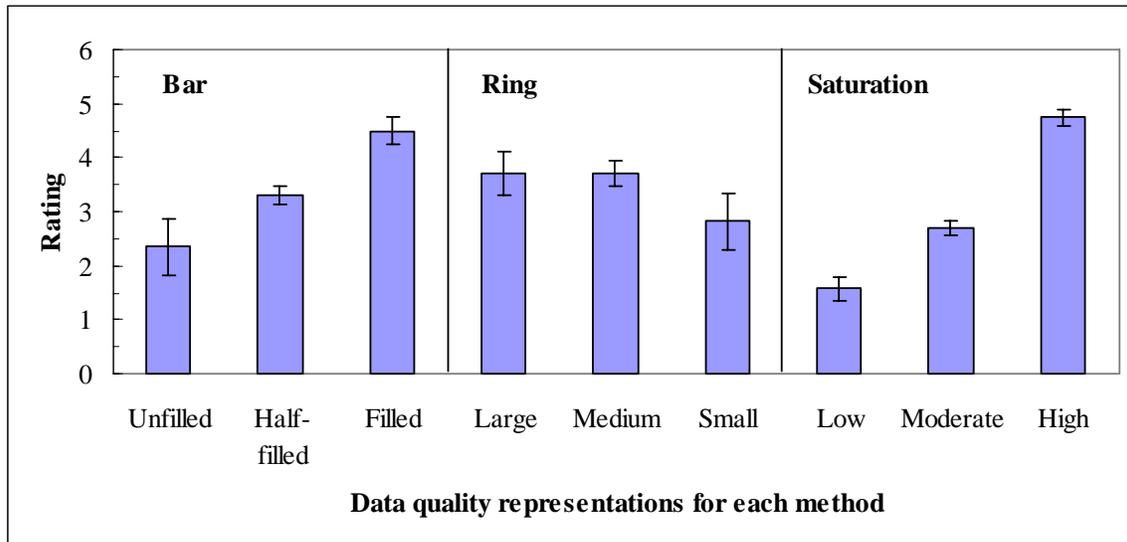

Figure 4: Average rating for each representation in the three different methods. A higher rating implies better data quality.

The analysis of symbol shape with the 2525B symbols showed no effect of symbol shape or interaction between symbol shape and representation for any of the methods. With the NTDS symbols, there was a significant effect of symbol shape on ratings with the saturation method ($F(8,56) = 4.1$), but no significant differences were found in the post-hoc analysis.

As indicated in Figure 5, response time did not vary across the different methods or across the different representations in the bar and ring methods. However, in the saturation method, the response time did vary as a function of representation ($F(2,14) = 26.4$) with the time to find and rate the saturated symbol being significant faster than with the other two representations.

The analysis of symbol shape showed a significant effect of symbol shape on response time for all three methods (bar: $F(8,56) = 27.3$, ring: $F(8,56) = 24.2$, saturation: $F(8,56) = 15.0$) with the NTDS symbols. There was also a small but significant interaction between symbol shape and representation for the ring method ($F(16,112) = 2.3$). With the 2525B symbols, there was a significant effect of symbol shape ($F(8,56) = 12.5$) and a significant interaction between symbol and representation ($F(16,112) = 12.5$) on response time with the saturation method only. The effects for symbol were similar to what was seen in Experiment 1; namely, people located the

surface symbols faster than the other symbols. With the ring method, response times for the NTDS surface symbols tended to increase as the diameter of the ring decreased. With the saturation method, the response times increased as the saturation decreased for all the symbols except the surface friendly for both symbol sets. However, the interaction was only significant for the 2525B symbol set.
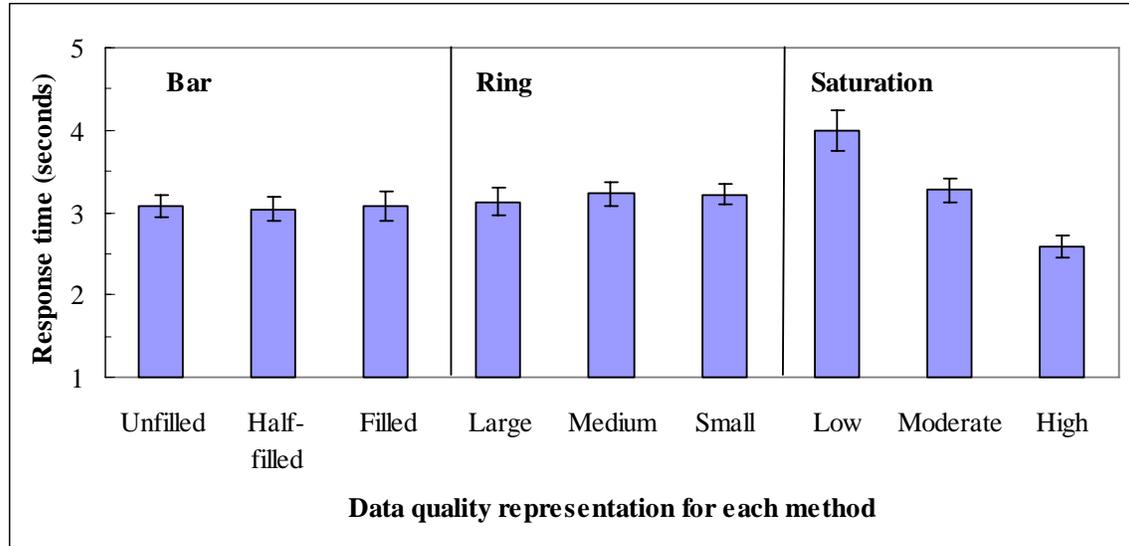

Figure 5: Average response time for each representation in the three different methods.

Discussion

The current experiments examined the impact of three different methods for representing data quality or uncertainty on a tactical display. The optimum method will lead to a consistent and intuitive interpretation of the quality of the data represented by the tactical symbol and will not interfere with the interpretability or visibility of the primary symbology. As indicated by the overall results from the two experiments, none of the methods achieved this goal. The saturation method resulted in the most consistent response, but it severely impacted the detection of the tactical symbols. The bar method had the least negative impact on detection, but there was some inconsistency in the participants' interpretation of the different levels of fill and response times were slower than in the control condition. The rings were the least successful. They were not consistently interpreted across participants and they interfered with detection.

The results for the saturation method were not unexpected. In general, colour coding leads to improved search times (Christ 1975). With the saturation method, people had to search for a symbol that could be one of three colours instead of looking for a specific colour. Thus, they probably had to put greater reliance on symbol shape. Evidence for this is the similarity between the current results and those in a related study (McFadden et al. 2007) using monochrome versions of the NTDS and 2525B symbols. In both studies, performance with the surface symbols, especially the surface friendly symbol, was significantly more accurate and faster than with the air and subsurface symbols. The search times for the saturation method in the rating task provide additional evidence. Response time increased as the target became less saturated.

However, part of this increase in response times could have been an artefact of the experimental design. People were always shown an example of the saturated symbol at the beginning of the run. As well, response times were a function of both the time taken to find the symbol and the time needed to rate the data quality. Thus, part of the increase in response time could be because it was more difficult to make a decision with the less saturated symbols. However, the rating results tend to discredit that hypothesis.

There is another potential problem with the use of saturation coding. In this study, considerable care was taken to make sure the three different saturation levels were discriminable. Room lighting was carefully controlled and participants could not adjust the display contrast and brightness. Discrimination of the different saturation levels can be adversely affected by the presence of glare on the screen or by large changes in the gain (contrast control) and offset (brightness control) of the display. Thus, the consistency with which the saturation coding was interpreted could decrease substantially in an operational setting.

The rating results for the rings could be due to the use of non-military participants. Range rings are often used to provide an estimate of the uncertainty associated with the position of a target and the larger the ring the greater the uncertainty. Some of the participants saw it that way while other probably equated bigger with better. Thus, the results might be different if a Navy population were used. On the other hand, the rings used on tactical displays are much larger as they are intended to indicate the area where the target could be. In this case, the different diameters of rings were associated with three different categories of data quality and the differences in the diameters of the rings were relatively small. The task was somewhat different. The issue was not where a track could be, but how accurate is the information represented by the symbol. Thus even with Navy personnel, the concept of bigger is better might be a factor. There were other problems with the rings. Response times in the counting task were significantly poorer with the rings compared to the control condition especially for the NTDS surface friendly symbol which is usually the most conspicuous. The presence of the ring could have made it more similar in appearance to the other symbols.

The bars had the least negative impact on the counting task. Accuracy was similar to the control condition although there was still a significant decrement in response time. This would suggest that adding a secondary symbol similar in footprint to the bar will not interfere substantially with the visibility of the tactical symbols. The interpretation of the bar was not as consistent as in the previous study (Unger Campbell and Baker 2003). Although none of the participants saw the unfilled as representing best data quality and the filled poorest, some did give a similar rating to all three levels of fill. Since there were always examples of the three levels of fill on the display on each trial, difficulty in discrimination should have shown up as a slower response time. Moreover, an examination of the variance suggests that people were responding consistently. Thus, it is more likely that some people did not intuitively associate the different levels with different levels of data quality. Since this was not a problem with the saturation method, it suggests that the different forms of the bar were not as intuitive as different degrees of saturation.

Experiments 3 and 4

Introduction

Given the pattern of results in Experiments 1 and 2, it would appear that the most effective representation would be one that combines a symbol such as the bar or ring with saturation. To assess this proposal, the ring and bar conditions were repeated using bars and rings that were either monochrome, as in Experiment 1 and 2, or varied in saturation. In the saturation condition, the filled bar and the smallest ring were the same colour as the symbol they were associated with; the half-filled bar and middle size ring were the same colour but less saturated than the symbol colour; and the unfilled bar and largest ring were even more desaturated. The presence or absence of redundant saturation coding of the bars and rings was a between participant factor. Given our findings in the previous experiments, it was hypothesized that people in the redundant saturation condition would more consistently associate the different variants of the bar and ring with different levels of data quality compared to people in the monochrome condition. Both the search task and the rating task were carried out to assess whether the use of variably coloured rings and bars had a different effect on search than their monochrome equivalent. Since no interaction was found between symbol set and method in the previous experiments and the interactions between symbol shape and method primarily occurred with the NTDS symbols, only that symbol set was tested.

Method

Participants

Twelve participants, 6 males and 6 females, took part in the two experiments. They ranged in age from 19 to 51 with a mean age of 28. All participants had normal or corrected-to-normal vision based on self-report and normal colour vision as assessed by the Ishihara pseudoisochromatic plates.

Conditions

Experiment 3 (counting task) was a 2 (colour coding of data quality symbols – grey or saturation) by 3 (method of representing data quality – none, bar, rings) design. The presence or absence of saturation coding was a between participant factor and data quality method a within participant factor. The design for Experiment 4 was identical except that there was no control condition. Participants carried out 27 runs (3 methods by 9 symbol types) in Experiment 3 and 18 runs (2 methods by 9 symbol types) in Experiment 4. The order in which the three methods were carried out was randomized across participants.
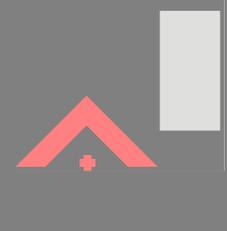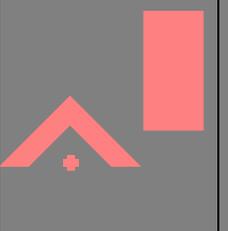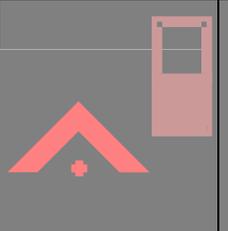
Apparatus

The apparatus was identical to the first two experiments.

Stimuli

The grey data quality symbols were identical to the previous experiments. In the colour coded conditions, the chromaticity coordinates shown in Table 4 were used to colour the three forms of the bar and ring. Thus, the filled bar and small ring had the same colour coding as the tactical symbol they were associated with, the half-filled bar and middle ring were coded with the category B colours and the unfilled bar and large ring with category C. Examples of the different data quality representations for the bar and ring methods are shown in Table 7.

Table 7: Examples of each data quality representation for the two different versions of the bar and ring methods with an air hostile tactical symbol.

Level of Uncertainty	Method of Data Quality Representation			
	Bars	Rings	Bars + saturation	Rings + saturation
A				
B				
C				

Display and tasks

The display and tasks were identical to the first two experiments.

Procedure

The procedure was relatively similar to the previous experiments. Since only one symbol set was used and only two methods were compared, the experiments required only 2 sessions of about 90 minutes each (including breaks). Experiment 3 was completed in the first session and Experiment 4 in session 2.

Results

Experiment 3 – counting task

As in the Experiment 1, the results of interest were counting accuracy and response time per trial for correct responses. The effect of group, method and symbol type on these three measures were analysed using a mixed between/within design. Main effects and interactions were included in the model and the Scheffe test was used for post-hoc analysis. A multivariate analysis of the over-all effect of accuracy and response time was also carried out.

As shown in Figure 6, participants in the redundant saturation condition tended to be more accurate but slower. However, there was no significant effect of group on accuracy, or response time, or a significant interaction between group and either method or symbol type. There was a significant effect of method and symbol type on accuracy and response time as well as an overall effect of the two measures (Table 8). A post hoc analysis, indicated that the bar and control methods were not significantly different but that performance was significantly less accurate with the ring method. With response times, there was also a significant difference between the control and the bar method. The results for symbol type were similar to those found in Experiment 1. Performance with the surface symbols tended to be more accurate and faster except for the ring method which impacted negatively on response times for the surface friendly symbol.

Table 8: Summary of within participant analyses of variance for accuracy, response time, and efficiency

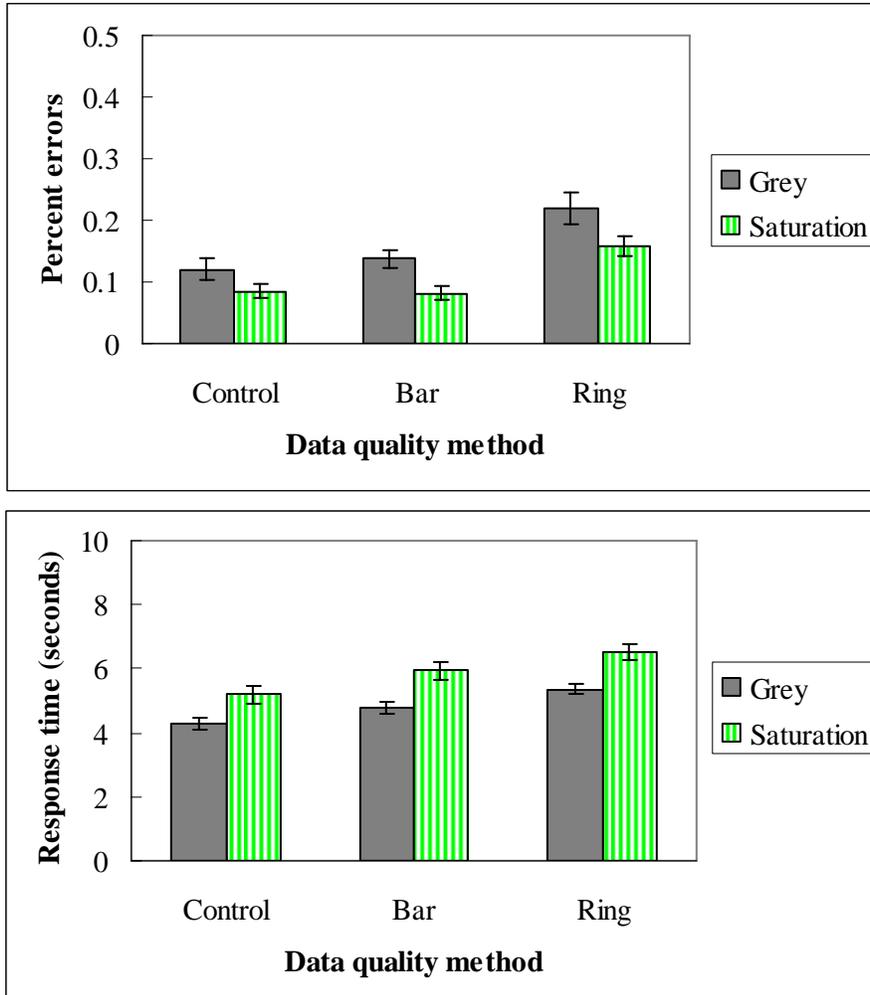| Source	Univariate / multivariate degrees of freedom	F values (P <0.01)		
		Accuracy	**Response time**	**Multi-variate**
Method	2,20 / 4,38	14.5	41.5	20.5
Symbol	8,80 / 16,158	8.6	159.3	32.0
Method by symbol	16, 160 / 32,318	n.s.	9.3	4.8

Figure 6: Effect of data quality method and colour coding of data quality symbol on accuracy and response time.

Experiment 4 – rating task

The results for Experiment 4 followed a similar pattern to those of Experiment 2. As shown in Figure 7, participants responses were more consistent with the bar method than with the ring method. Moreover, this pattern did not change with the addition of the saturation coding. As with the result for Experiment 2, a separate analysis of the effect of symbol type, and representation on rating response was carried out for each method. A similar analysis was carried out on the median log response times. In addition, the effect of method on response time was examined.
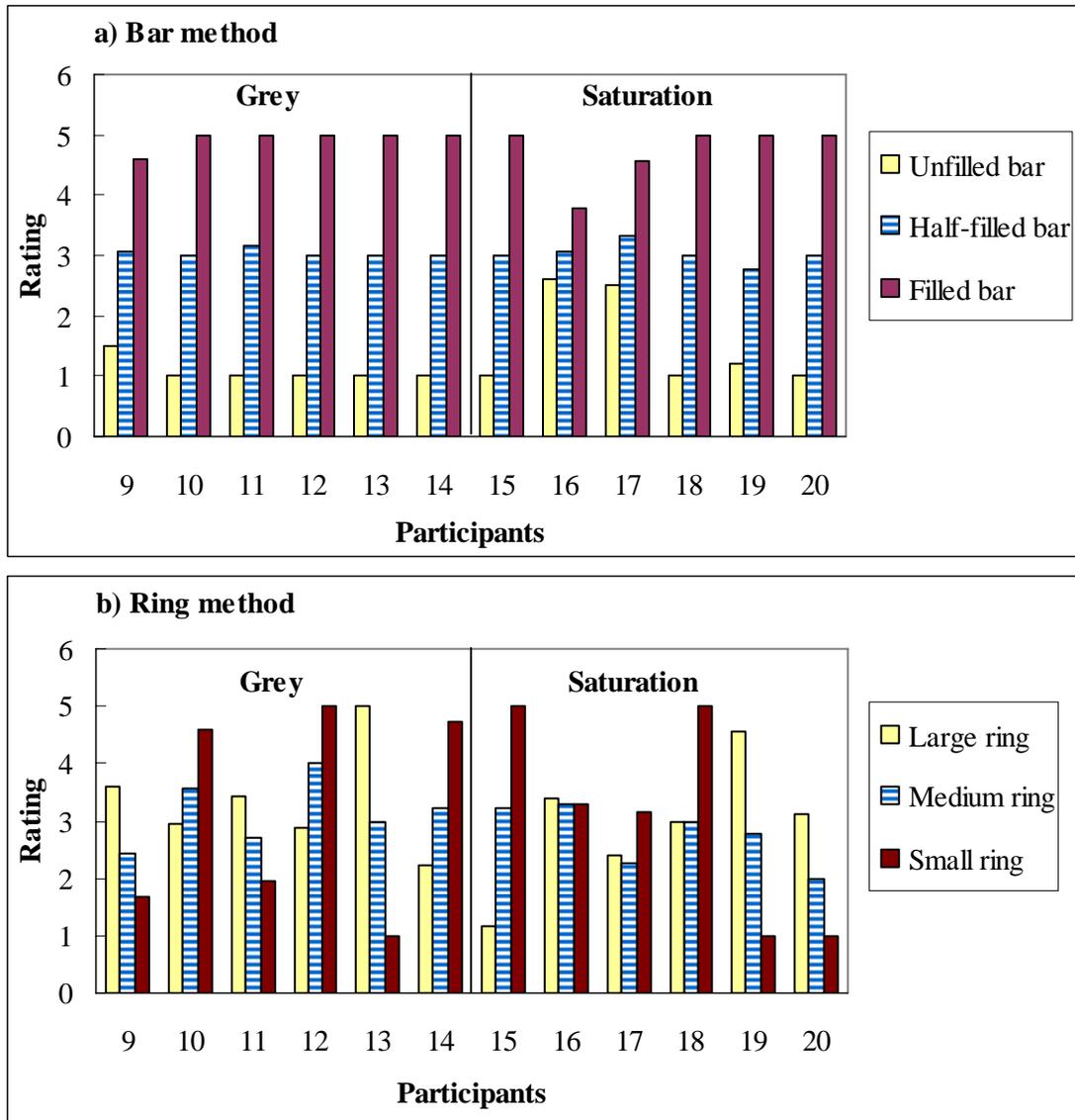
Figure 7: Average rating by each participant for each representation for data quality method. Ratings were always between 1 and 5 and a higher rating indicated greater certainty.

Again, there was no effect of group and no interaction between group and either symbol type or representation on the ratings. Ratings for each representation were significantly different for the bar ($F(2,20) = 173.8$, $p < 0.01$) method, but not for the ring method (Figure 8a). The post-hoc analysis showed that the rating for each representation was significantly different from the remaining two in the bar method with the filled bar being associate with high data quality and the unfilled bar with poor data quality. There was no effect of symbol type for either method.

Unlike Experiment 2, method did have an effect on response time. Participant responded significantly faster with the bar method than with the ring method ($F(1,10) = 45.7$) (Figure 8b).
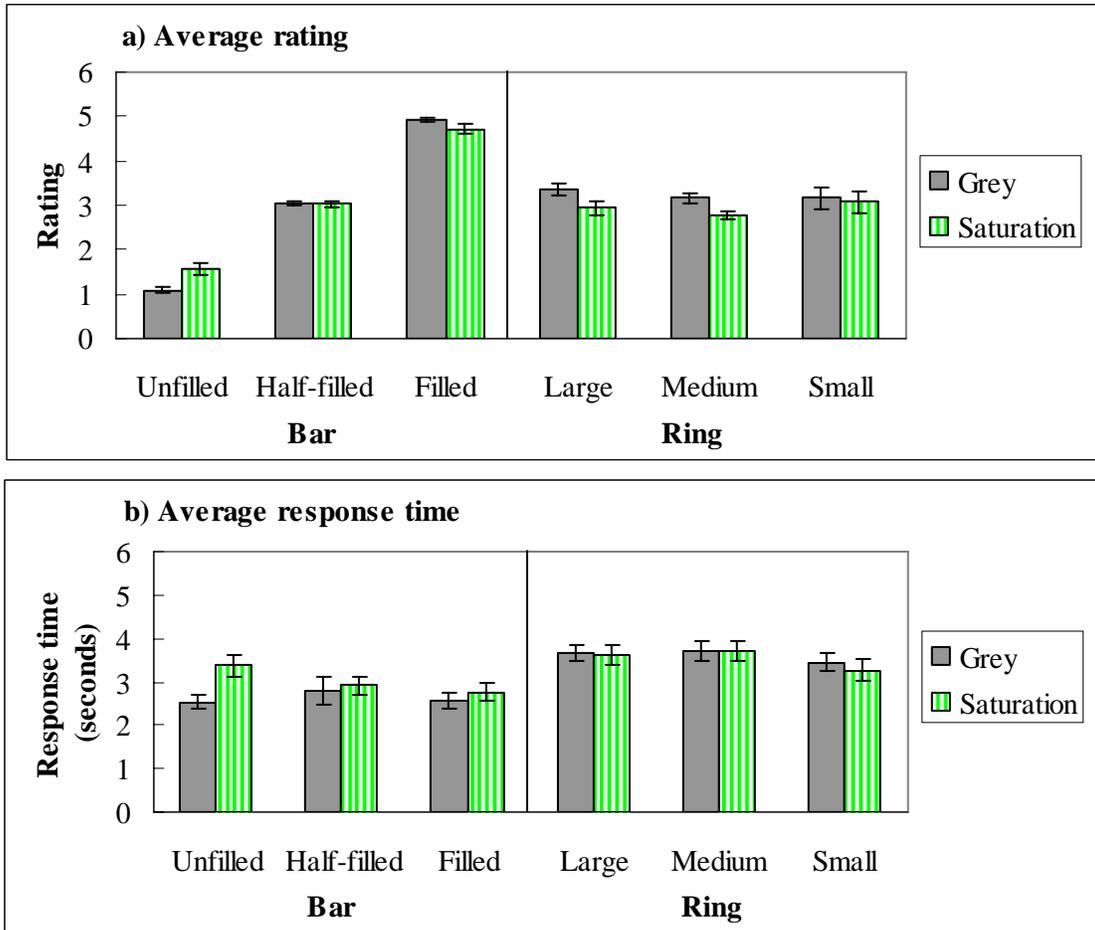
Figure 8: Average rating (a) and response time (b) for each representation for the two groups and methods.

Discussion

The addition of saturation coding appeared to have little impact on either search performance or response consistency. In terms of search performance, the results were similar to Experiment 1. The bar interfered less with counting performance than the ring although response times were slower than in the control condition. Unlike the previous experiment, all the participants interpreted the bar in the same way associating the filled bar with high data quality and the unfilled bar with low data quality. Interpretation of the different sizes of rings still varied across observers even when saturation was used as a redundant code. Some participants associate the largest ring with highest certainty and some rated the smallest ring that way. The instructions could have contributed to the lack of an effect. Participants in the two groups were given identical instructions. Those in the colour coding group may have ignored the saturation of the data quality symbol. As well, the rings are not very thick on the display at the standard viewing angle. The different levels of saturation coding may not have been as visible as they were when the saturation of the tactical symbols themselves was changed.

General discussion

As operators of command and controls systems become more removed from the source of the data underlying the information they are using, it is more difficult for them to ascertain the accuracy of that information. This applies not only to algorithmically generated information but also to information sent by allies, other platforms etc. For this reason, information associated with contacts or tracks often includes an assessment of the quality of the underlying data. An operator can access the underlying information by hooking the track. In a cluttered display, this could be a time consuming task and it would be very difficult for the operator to maintain a global picture of the relative quality of a large number of tracks. Without this understanding, the tendency is either to perceive the displayed information as very accurate or to ignore it. It was this concern that led to the current research under COMDAT.

In the previous study (Unger Campbell and Baker 2003), operators expressed their concern about clutter and about increasing their secondary task load by requiring them to interpret the data quality symbology. Thus, they favoured the least intrusive data quality symbology tested. Of the four symbol sets evaluated, the slider (the bar in this study) was most intuitive in that operators interpreted the different levels consistently.

The experiments reported in this paper had two goals. The first was to address the operators' concern about clutter. The second was to explore some alternative concepts for representing data quality that did not have the same perceived limitation of increasing clutter. Three methods were evaluated – adding a separate symbol (the bar), annotating the symbol by adding different diameter rings, and modifying the symbol by varying the colour saturation. The counting task was used to assess the effect of adding a representation of data quality on the operator's ability to detect tracks quickly and accurately. The rating task assessed the intuitiveness of the proposed methods.

The overall finding was that the use of a separate symbol such as a slider or bar interfered least with the rate and accuracy with which tactical symbols were located. Counting performance was as accurate but somewhat slower relative to a control condition in which there was no representation of data quality. Moreover, the majority of the participants interpreted the bar symbol in the same way. A small number of participants did not see the different representations of the bar as representing different levels of data quality, but none of them interpreted the different representations in the opposite direction. Thus a small symbol, with a footprint equal to or smaller than that of the bar symbol, should be a reasonable method for representing data quality providing no more than three levels are required. If a different type of symbol is considered it would be important to assess the consistency with which it is interpreted. However, it is important to provide the ability to remove the data quality representation since it could impact the time taken to locate critical symbols.

Saturation coding was the most consistently interpreted in the rating experiment. However, it had the greatest negative impact on search performance. It has the additional problem of being susceptible to changes in ambient illumination and monitor settings (e.g. the offset or brightness control). Thus, the consistency with which the saturation coding was interpreted could decrease substantially in an operational setting. Combining saturation coding with the three symbols did

not impact the consistency with which people interpreted the different representations. Overall, saturation coding should probably be avoided.

The rings were not consistently interpreted and search performance was poorer than with the bars. Overall, the results with the ring indicate that it is important to evaluate proposed methods for representing data quality with a substantive user group. This recommendation is also supported by Unger Campbell and Baker (2003) study and the results from Finger and Bisantz (2002). In addition, the current results suggest that there is probably a limit to the amount of information that can be incorporated into a symbol. The different dimensions may interact in unexpected ways. In this case, the rings appear to have reduced the conspicuity of the NTDS symbol shapes.

One factor that was not addressed in this study was the type of data quality being represented. In the Unger Campbell and Baker study, the slider tended to be associated with time lateness or the time since the most recent update. In the current study, no specific type of data quality was specified although people were given examples of possible types. In its most basic form, a track symbol provides location, platform, and ID information. Predicted direction can also be displayed. An MSDF system may associate a probability with each of these dimensions. An attempt to represent all of these dimensions in one symbol was rejected by participants in the Unger Campbell and Baker study as being too intrusive and also as difficult to interpret. Thus, it is probably better if the symbol represents either the most critical dimension for the operator or a general statement about the overall quality of the information associated with that track. In the former case, it would be necessary to provide operators with the ability to specify the type of data quality that they wished to see represented.

In these experiments and the one by Unger Campbell and Baker, only three different representation of each symbol were presented. This decision was based on human factors guidelines for symbolic coding that recommends the use of only a small number of levels along any one coding dimension. Other studies ((Banbury et al. 1998, Finger and Bisantz 2002)) have used many more levels. The results of those studies suggest that people tend to group these multiple levels into two or three categories although the boundary between successive categories is not that well defined. This trend would tend to support our decision to restrict the number of levels to three. On the other hand, the Navy, in its classification process, uses four levels - Possible Low (Poss Low), Possible High (Poss High), Probable (Prob), and Certain (Cert) to categorize confidence in a classification decision. A representation of data quality that corresponded to these four levels would probably receive greater acceptance. However, it would be necessary to determine if an intuitive symbol could be designed that showed more than three clearly discriminable levels.

The study by Finger and Bisantz (2002) also looked at the value of including a numerical estimate of uncertainty or confidence along with the symbolic representation. As stated earlier, people tended to be more conservative and more accurate in their identification of whether a target was hostile or friendly when a numerical probability was included. Thus, if more detailed representation of data quality is required, a numeric representation might be more suitable. Research would be required to determine how to group the underlying probabilities to provide an intuitive mapping between the precise probability and the displayed number or word. Further research is clearly needed on the number of levels that can be effectively utilized and the effect of symbolic versus alphanumeric methods for representing data quality.

Recommendations

Based on the existing literature and the results of this study, a small monochrome symbol such as the bar or the slider used in these studies should provide an operator with an intuitive estimate of overall data quality provided by a tactical symbol and should not interfere with the operator's ability to find the tactical symbols. Nevertheless, operators should have the option of turning the supplementary symbol off in case they felt that the additional symbology made the display too cluttered. If more detailed information is required then an alphanumeric representation should be used.

Further research is required to determine how many data quality levels operators can utilize and whether there is an intuitive mapping of actual probabilities onto these levels. In addition, further research is required to determine the effective use of alphanumeric levels or categories for representing estimates of data quality as an addition or replacement for symbolic representations.

Conclusion

Multi-Source Data Fusion (MSDF), as developed under COMDAT, provides an assessment of the reliability of the estimate of the fused track's attributes and position. Part of the human factors work under COMDAT has been to investigate methods of representing the quality of the MSDF-generated tracks to the operator. The research reported in this paper was concerned with the potential effect of different representations of data quality or uncertainty on the visibility of tactical symbols and the intuitiveness of the different representations. Three methods of representing data quality were investigated: a variably filled bar presented beside the track symbol, different diameter rings that surrounded the tactical symbol, and varying the saturation of the tactical symbol itself. In Experiment 1, a visual search task was used to compare the accuracy and speed with which participants could locate multiple instances of each of the tactical symbols without any representation of data quality and with each of the three methods tested. Experiment 2 examined the ability of participants to quickly and consistently interpret the data quality represented by the tactical symbol as a function of the three different methods. The results indicated that the bar interfered least with people's ability to locate the tactical symbols, but the saturation method was most consistently interpreted. The rings degraded detection of the tactical symbols and were not consistently interpreted. A second set of experiments looked at whether using the saturation method to redundantly code the bars and rings, would improve the consistency with which they were interpreted without negatively affecting the visibility of the tactical symbols. Redundant saturation coding did not affect response consistency. However, participants in this study all tended to consistently interpret the bars. The basic recommendation was that a small independent symbol such as the bar was the preferred method for representing data quality. However, operators should have the option of removing the data quality symbology to reduce clutter. If a different symbol shape is chosen, its intuitiveness should be assessed prior to implementation. Further research is required to improve our understanding of the number of categories of data quality operators can use effectively.

References

Banbury, S., Selcon, S., Endsley, M., Gordon, T., and Tatlock, K. (1998). Being certain about uncertainty: How the representation of system reliability affects pilot decision making, In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*. Vol. 1, 36-39, Santa Monica, CA: Human Factors and Ergonomics Society.

Christ, R. E. (1975). Review and analysis of color coding research for visual displays. *Human Factors*, 17, 542-570.

Department of Defence (1999). Common warfighting symbology (MIL-STD-2525B). Department of Defense, Washington, DC.

Edwards, L. D. and Nelson, E. S. (2001). Visualizing data uncertainty: a case study using graduated circle maps. *Cartographic Perspectives*, 38, 19-36.

Evans, B. J. (1997). Dynamic display of spatial data-reliability: does it benefit the map user? *Computers & Geosciences*, 23 (4), 409-422.

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

Harrower, M. (2003). Representing uncertainty: does it help people make better decisions?, In *UCGIS Workshop: Geospatial Visualization and Knowledge Discovery Workshop*. National Conference Centre, Landsdowne, VA.

Howes, A. and de Bruijn, O. (2005). A review of approaches to displaying uncertain and complex information. School of Informatics, University of Manchester, Manchester, UK. (Draft report).

Hutton, R. J. B. (2004). Types of "uncertainty about" and strategies for uncertainty management: an integrative approach. (#04TA2-SP1-RT2). Klein Associates Inc., Fairborn, OH.

Kirschenbaum, S. S. and Arruda, J. E. (1994). Effects of graphic and verbal probability information on command decision making. *Human Factors*, 36 (3), 406-418.

Lockheed Martin Canada Inc. (2001). COMDAT uncertainty trade off analysis. Lockheed Martin Canada Inc., Montreal, PQ. (Power Point Presentation).

McFadden, S. M., Jeon, J., Li, A., and Minniti, A. (2007). Evaluation of symbol sets for naval tactical displays. (DRDC Toronto TR 2007-046). Defence R&D Canada - Toronto, Toronto, ON.

Unger Campbell, G. and Baker, K. (2003). Command Decision Aiding Technology (COMDAT) symbology and design study. (DRDC Toronto CR 2003-036). Unger Campbell and Associates, Langley, BC.

Unger Campbell, G. (2004). Command Decision Aiding Technology (COMDAT) Operator-Machine Interface (OMI) style guide: Version 2.0. Unger Campbell and Associates, Delta, B.C.

Waldron, S. M., Duggan, G. B., Patrick, J., Banbury, S., and Howes, A. (2005). Adaptive information fusion for situation awareness in the cockpit, In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*. 49-53, Santa Monica, CA: Human Factors and Ergonomics Society.

This page intentionally left blank.

Annex A   Generation of different saturation levels

The different saturation levels used in these experiments were generated using a modified version of a method described by De Spirito (2002). The method is based on the assumption that desaturated colours are colours that are mixed with grey. Thus it is possible to generate desaturated versions of a colour on an electronic display by adding different amounts of grey to it. The formula shown in the article was as follows:

Table A1: Original formula for calculating RGB values for desaturated colours

RGB	Orange	Grey		50% Orange
Red	255	+ 128	/ 2	= 192
Green	128	+ 128	/ 2	= 128
Blue	0	+ 128	/ 2	= 94

It produces a 50% saturated orange of equivalent luminosity according to the author. More desaturated colours can be produced by repeating the calculation, but substituting the new RGB values in column two. No general formula was provided nor did the author point out that in the general case the luminosity (or sum of the RGB values) of the grey and the original colour should be the same.

In order for this method to work for all colours, it is necessary to either normalize the new RGB values so that they sum to the same amount as the original values or to choose a grey whose RGB values sum to the same amount as the original colour. We used the first method. The example below shows the modified formulae for the blue used in our study to create a blue with 50% saturation. Again, other saturation levels could be achieved by substituting the new RGB values in column 2.

Table A2: Modified formula for calculating RGB values for desaturated colours

RGB	Blue	Grey		50% blue		50% blue at original luminance
Red	255	+ 128	/ 2	= 192	/ (638/511)	239
Green	255	+ 128	/ 2	= 192	/ (638/511)	239
Blue	128	+ 128	/ 2	= 128	/ (638/511)	160
Total	638			511		638

The above formula was used to create different saturation levels for each of the three original colours. Based on visual inspection, saturation levels of 33% and 11% were chosen as being reasonably discriminable from each other and from the original colour.

Since RGB values are used, the chromaticity coordinates and luminance of the resulting colours will be different on different monitors. This was not a problem for the current study. Since we were only concerned that the saturation levels be clearly discriminable under the conditions we were testing. Moreover, MIL-STD-2525B does not specify the chromaticity coordinates for the colours used for affiliation coding. Thus, a method based on the symbols RGB values is more reasonable as it insures internal consistency across monitors.

The RGB values calculated using the above method are shown in the following table along with the chromaticity coordinates reported in the body of the paper.

Table A3: Monitor RGB values and measured CIE 1931 chromaticity coordinate and luminance for 9 colours used in saturation condition.

Saturation level	Units	Affiliation		
		Friendly	Hostile	Unknown
100%	x, y, l	0.211, 0.262, 66	0.411, 0.326, 40	0.358, 0.442, 92
	RGB	128, 224, 255	255, 128, 128	255, 255, 128
33%	x, y, l	0.241, 0.283, 62	0.325, 0.313, 39	0.315, 0.373, 78
	RGB	170, 212, 226	204, 153,153	232, 232,174
11.1%	x, y, l	0.263, 0.299, 61	0.291, 0.308, 40	0.291, 0.331, 70
	RGB	190, 206, 211	182, 164, 164	220, 220, 198

References

De Spirito, E. (2002). Understanding color and the RGB system (online), Latium Software, www.latiumsoftware.com/en/articles/00015.php. (June, 2007).

List of symbols/abbreviations/acronyms/initialisms

CCS	Command and Control System
COMDAT	COMmand Decision Aiding Technology
DRDC	Defence Research and Development Canada
HMCCS	Halifax Class Modernization Command and Control System
HREC	Human Research Ethics Committee
MIL-STD-2525B	Military Standard 2525B
MSDF	Multi Source Data Fusion
NTDS	Naval Tactical Data System
OMI	Operator-Machine Interface
SME	Subject Matter Expert
TDP	Technology Demonstrator Project

This page intentionally left blank.

Distribution list

Document No.: DRDC Toronto TR 2007-032

LIST PART 1: Internal Distribution by Centre:

2 DRDC Toronto Library file copies

5 Sharon McFadden

7 TOTAL LIST PART 1

LIST PART 2: External Distribution by DRDKIM

1 Defence R&D Canada

1 DRDKIM

1 DSTM

   National Defence Headquarters
   Louis St Laurent Bldg
   555 boul. de la carriere
   Gatineau, QC K1A 0K2

1 Maritime Research Co-ordination Cell

1 DMRS 8

1 DMSS 8

1 PM/HCM/FELEX

   Canadian Forces Experimentation Centre (CFEC)
   3701 Carling Avenue
   Ottawa ON K1A 0Z4

1 CFEC/P. Labbé

   DRDC Atlantic
   9 Grove St
   Dartmouth NS B2Y3Z7

1 Dr. Bruce McArthur

1 Mr. Nelson McCoy

1   Dr. Bruce Chalmers

1   Dr. Jacquelyn Crebolder

1   Mrs. Liesa Lapinski

     Defence R&D Canada - Valcartier
     2459 Pie Xi Blvd North
     Val-Belair, QC G3J 1X5

1   Mr. Stéphane Paradis

1   Mr. Eloi Bossé (Thrust Leader 1b)

1   PM/INCOMMANDS

15   TOTAL LIST PART 2

22   TOTAL COPIES REQUIRED

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

Multi-Source Data Fusion (MSDF), as developed under COMDAT, provides an assessment of the reliability of the estimate of the fused track's attributes and position. Part of the human factors work under COMDAT has been to investigate methods of representing the quality of the MSDF-generated tracks to the operator. The research reported in this paper is concerned with the potential impact of different representations of data quality or uncertainty on the visibility of tactical symbols and the intuitiveness of the different representations. Three methods of representing data quality were investigated: a variably filled bar presented beside the tactical symbols, different diameter rings that encompassed the tactical symbols, and varying the saturation of the tactical symbols. In Experiment 1, a visual search task was used to compare the accuracy and speed with which participants could locate multiple instances of each of the tactical symbols without any representation of data quality and when each of the three methods tested were added. Experiment 2 examined the ability of participants to quickly and consistently interpret the quality of the data, represented by the tactical symbol, using the three different methods. The results indicated that the bar interfered least with people's ability to locate the tactical symbols, but the saturation method was most consistently interpreted. A second set of experiments looked at applying the saturation coding to the bars and rings instead of the tactical symbol. Redundant saturation coding of the bars and rings had no effect on tactical symbol visibility and did not improve the consistency of interpretation of data quality. The basic recommendation was that a small independent symbol, such as the bar, was the preferred method for representing data quality. However, operators should have the option of turning it off if the display appears too cluttered. If a different symbol shape is chosen, its intuitiveness should be assessed prior to implementation. Further research is required to improve our understanding of the number of levels of data quality operators can use effectively.

La fusion de données de multiples sources (Multi-Source *Data Fusion,* MSDF), telle que mise au point dans le cadre du COMDAT, fournit une évaluation de la fiabilité de l'estimation des attributs et de la position des trajectoires. Les travaux sur les facteurs humains menés dans le cadre du COMDAT ont en partie consisté en l'étude de méthodes de représentation pour l'opérateur de la qualité des trajectoires générées par la MSDF. Les travaux de recherche dont il est fait état dans le présent rapport abordent l'incidence potentielle de différentes représentations de la qualité des données ou l'incertitude rattachée à la visibilité des symboles tactiques et le caractère intuitif de différentes représentations. Trois méthodes de représentation de la qualité des données ont été étudiées : une barre plus ou moins remplie placée à côté des symboles tactiques, des anneaux de différents diamètres autour des symboles tactiques et la variation de la saturation des symboles tactiques. Dans la première expérience, une tâche de recherche visuelle a été utilisée pour comparer l'exactitude et la vitesse avec lesquelles les participants pouvaient localiser de multiples occurrences de chacun des symboles tactiques sans représentation de la qualité des données et dans le cas de chacune des trois méthodes éprouvées. Dans la deuxième expérience, on a examiné l'aptitude des participants à interpréter rapidement et de manière uniforme la qualité des données, représentées par le symbole tactique, pour chacune des trois méthodes différentes. Les résultats indiquent que c'est la barre qui nuit le moins à l'aptitude des gens à localiser les symboles tactiques, mais la méthode de la saturation est celle qui permet l'interprétation la plus uniforme. Dans un deuxième ensemble d'expériences, on a examiné l'application d'un codage en saturation des barres et des anneaux plutôt que des symboles tactiques. Le codage redondant en saturation des barres et des anneaux n'a aucun effet sur la visibilité des symboles tactiques et n'améliore aucunement l'uniformité de l'interprétation de la qualité des données. La recommandation fondamentale est de préférer un petit symbole

indépendant comme la barre pour la représentation de la qualité des données. Cependant, les opérateurs devraient avoir le choix de le supprimer de l'affichage si celui-ci devenait trop encombré. Si un symbole de forme différente était retenu, il faudrait en évaluer le caractère intuitif avant la mise en œuvre. D'autres recherches sont nécessaires pour améliorer notre compréhension du nombre de niveaux de qualité des données que les opérateurs peuvent efficacement utiliser.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

symbology; uncertainty; Navy; tactical displays; visual search; classification; data quality; operator-machine interface

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE **R&D** DÉFENSE

www.drdc-rddc.gc.ca