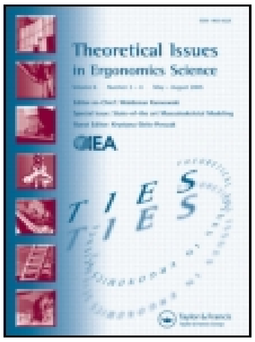

Taylor & Francis
Taylor & Francis Group

Theoretical Issues in Ergonomics Science

ISSN: 1463-922X (Print) 1464-536X (Online) Journal homepage: https://www.tandfonline.com/loi/ttie20

Utilizing graphical formats to convey uncertainty in a decision-making task

Richard Finger & Ann M. Bisantz

To cite this article: Richard Finger & Ann M. Bisantz (2002) Utilizing graphical formats to convey uncertainty in a decision-making task, Theoretical Issues in Ergonomics Science, 3:1, 1-25, DOI: 10.1080/14639220110110324

To link to this article: https://doi.org/10.1080/14639220110110324

Published online: 26 Nov 2010.

THEOR. ISSUES IN ERGON. SCI., 2002, VOL. 3, NO. 1, 1–25

Utilizing graphical formats to convey uncertainty in a decision-making task

RICHARD FINGER and ANN M. BISANTZ*

University of Buffalo, State University of New York, Amherst, NY 14260-2050, USA

Keywords: Uncertainty display; Decision-making.

Understanding how to display effectively uncertain information has become increasingly important as decision aids can provide operators with situational estimates and their associated uncertainty. The paper describes two studies in which degraded or blended icons were used to convey uncertainty regarding the identity of a radar contact as hostile or friendly. A classification study first showed that participants could sort, order and rank icons from five sets intended to represent different levels of uncertainty. Three icon sets were selected for further study in an experiment in which participants had to identify the status of contacts as either hostile or friendly. Contacts and probabilistic estimates of their identities were depicted on a simulated radar screen in one of three ways: with degraded icons and probabilities, with non-degraded icons and probabilities and with degraded icons only. Results showed that participants using displays with only degraded icons performed better on some measures and as well on other measures, than the other tested conditions. These results are significant because they indicate both that people can understand uncertainty conveyed through such a manner and thus that the use of distorted or degraded images may be a viable alternative to convey situational uncertainty.

1. Introduction

Understanding how effectively to display uncertain information has become increasingly important as information systems and decision aids can provide operators with situational estimates and their associated uncertainty. Uncertain or probabilistic information can be shown in a variety of formats ranging from simply text to graphical representations to text/graphical hybrids.

The question of how to represent or display uncertain information is important from two perspectives. First, is the theoretical need to determine the manner in which different representations or display formats may impact users' understanding of the uncertainty and thus subsequently affect the decisions and actions made based on the different representations. Second, and more practically, it is necessary to determine how best to display this information to users; particularly when there is uncertainty associated with a large number of objects or data points.

Prior research has provided information relevant to these questions to some extent. For instance, with respect to understanding possible effects of representation on users' understanding of uncertainty, we know from the extensive literature on subjective probability and human reasoning under uncertainty that people do not

*Author for correspondence at: 342 Bell Hall, Department of Industrial Engineering, University at Buffalo, State University of New York, Amherst, NY 14260-2050, USA. e-mail: bisantz@eng.buffalo.edu

Theoretical Issues in Ergonomics Science ISSN 1463–922X print/ISSN 1464–536X online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI 10.1080/1463922011011032 4

always reason in a manner that would be prescribed by normative theories of probability. In some cases, people exhibit consistent use of reasoning heuristics that result in predictable biases (Bar-Hillel 1980, Tversky and Kahneman 1982a–d). However, research has indicated that factors such as task content, experience and information format can affect the degree to which people deviate from normative reasoning (Gigerenzer 1994, Koehler 1996), across various contexts including medicine (Christensen-Szalanski and Bushyhead 1981, Christensen-Szalanski and Beach 1982) and auditing (Butt 1988, Solomom *et al.* 1985). For instance, some researchers have found that providing uncertain information in a frequency format (e.g. 1/10), rather than a probability format (e.g. 1) improves participants' use of base rates in estimating posterior probabilities (Gigerenzer and Hoffrage 1995, Cosmides and Tooby 1996).

There are some differences between research in subjective probabilities, which can be defined as a person's 'degrees of belief in the truth of particular propositions (Ayton and Wright 1994: 164)' and the information communication problem of displaying uncertainty regarding estimates produced by a decision aid. Unlike much of the research in subjective probability, the question is not of having people make judgements of probabilities, but rather presenting known probabilities to people such that this information can be incorporated into their subsequent decision processes. However, even in this latter case, the human decision-maker is assessing and thus internalizing a subjective notion of the uncertainty represented by the decision aid output. Therefore, one would expect that, as with assessments of subjective probabilities, decisions based on decision aid estimates may be impacted by the manner in which the information is presented.

Also relevant are the results of research investigating linguistic representations of uncertainty (e.g. words such as likely, probable and unlikely). Such representations can have been described as 'vague' representations of uncertainty, compared to precise representations such as numeric probability values (Wallsten 1990, Wallsten and Budescu 1995). Overall, this research has indicated that linguistic representations of probability tend to be interpreted vaguely by people, covering a broad range of probabilities, and that individual assignments of probability values to linguistic expressions can be modelled using fuzzy membership functions (Wallsten *et al.* 1986, for a review of this literature see Budescu and Wallsten 1995, Wallsten and Budescu 1995). However, there tend to be individual differences in the manner in which people both interpret and generate linguistic expressions of probability values (Budescu and Wallsten 1985). Additionally, researchers have found effects of task context on peoples' preference for and interpretation of linguistic probability phrases. For instance, Dusenbury and Fennema (1996) found that participants had a preference for numeric representations when a task was described in terms of a high risk gain, while they preferred linguistic representations when the task was described in terms of a high risk loss. Brun and Teigen (1988) found that problem context (e.g. current events, medical treatment, news reports) affected participants interpretation of linguistic probabilities. However, studies in which people had to make judgements regarding gambles or lotteries indicated little difference in judgement performance across linguistic and numeric representations of probability values (Budescu *et al.* 1988, Budescu and Wallsten 1990, Erev and Cohen 1990). Researchers have theorized that this may be due to the fact that, when making decisions, people treat interpret both numeric and linguistic representations of uncertainty in a vague way, but then convert these representations to point estimates of uncertainty to use in decision-making (Budescu and Wallsten 1995).

Thus, extensive research has indicated individual differences and context effects in the manner in which people interpret linguistic representations of uncertainty, but limited effects of linguistic versus numeric representations on decision-making tasks. However, this research has not been typically been extended to the study of graphical representations (but see Budescu *et al.* 1988) and has dealt primarily with judgement tasks involving lotteries or gambles.

There has, however, been some relatively limited research on the impact of graphical representations of uncertainty on decision-making in situations that are more realistic and dynamic than those typically used in the linguistic probability research (Kirschenbaum and Arruda 1994, Andre and Cutler 1998). In these cases, the nature of uncertainty regarded the locations of objects (typically, military-style entities), or their heading. For example, Andre and Cutler (1998) investigated the utility of rings to represent uncertainty regarding the location of an entity: the size of the ring was dependent on the level of uncertainty. Empirical evidence showed that, in a collision avoidance task, performance improved using such a display. Kirschenbaum and Arruda (1994) conducted a similar experiment that compared verbal and graphical representations of uncertain submarine locations. Participants were shown either an ellipse-shaped representation of the confidence interval of the submarine's location or a verbal indicator describing the quality of the indicated submarine location. Kirschenbaum and Arruda (1994) found some advantage in estimating distance to target for the graphical representation, for situations with greater uncertainty. Andre and Cutler (1998) also compared a numeric, arc and ring representation of uncertainty regarding an object's heading in a simulated anti-aircraft task and found that all three representations improved user performance in 'killing' hostile aircraft (which participants had to identify based on the object's heading) when compared with a no aid condition, under conditions of high uncertainty regarding heading. The arc-based aid, which represented the uncertainty in direction by using an arc that covered the entire angle of possible movement heading, provided a slight advantage over the other two aids.

These studies have provided a preliminary indication that graphical representations of uncertainty may provide value to decision-makers and that the impact may be greater under conditions of higher uncertainty. Thus, additional work in the area of graphical representations of uncertainty may prove fruitful. For instance, this research did not address questions such as how to represent uncertainty regarding the nature of the object itself (e.g. uncertainty as to what the object was), nor address concerns regarding representation of uncertainty in large amounts of data.

Within the scientific visualization and Geographical Information Systems communities, on the other hand, there has been work done to address the practical need to visualize or display the uncertainty inherent in large data sets. For example, Pang *et al.* (1997) addressed problems of visually displaying data and their associated uncertainty in a holistic fashion, to provide users with a more accurate understanding of the data. They describe techniques for displaying statistical uncertainty, errors or discrepancies between different estimates of the same data point and ranges of potential data values. Pang *et al.* (1997) described techniques for displaying uncertainties associated with data visualization itself. For instance, it is desirable to display to users the differences, or discrepancies, between visualizations of phenomenon generating using different algorithms or techniques; or between a visualization and the true values of the information. Pang *et al.* (1997) suggested modifying attributes of scene geometry to indicate uncertainty, such as using colour, shading, reflectivity and bumpiness. Animation can also be used, so that perceptual aspects of animation such as the degree of motion, correspond to the degree of uncertainty regarding a value. Similarly, researchers in geographic information systems have been concerned about the need to visualize data quality, such as the uncertainty related to a spatial location, or the variability associated with a characteristic of a geographic location (e.g. the variance of a population estimate of a census tract; MacEachren 1992). MacEachren (1992) argues that traditional graphic variables used in cartography (location, size, value, texture, colour, orientation and shape), with the addition of colour saturation, should be mapped onto different types of uncertainty in data. MacEachren (1992) suggests that representations of geographic data and its associated uncertainty, could be provided through pairs of graphics (i.e. one graphic showing the data value at a point in space, while the other graphic shows the uncertainty of the data at the corresponding point in space) either shown simultaneously or alternating, or through the use of texture and colour on the same graphic.

Additionally, Pang *et al.* (1997) suggest the use of glyphs (graphical forms such as arrows or vertical lines) to display data uncertainty. For example, Wittenbrink *et al.* (1996) used glyphs to represent magnitude and direction of winds and ocean currents along with the uncertainties in these dimensions. Uncertainty in this case stemmed from data samples taken over time and across multiple sensors. In one example, the general shape of the glyph was an arrow, with the width of the arrowhead representing uncertainty in heading and multiple arrowheads representing uncertainty in magnitude. In another example, the area of the glyph was used to represent magnitude. They term these representations 'verity visualizations.' Wittenbrink *et al.* (1996) claim that these glyph representations are improvements over other techniques for representing uncertainty, such as 'overloading,' or the use of the dimensions of colour or texture, for two reasons. First, the integrated glyph representation uses less 'ink' to represent information and its associated uncertainty than techniques that require a separate display or annotation of uncertainty. An increase in the 'data-ink ratio' (ratio of information displayed to ink used to display the information) has been proposed by Tufte (1983) as a metric for the quality of a data display. Second, providing uncertainty information directly in the glyph allows other graphical dimensions (colour, texture) to be utilized for other purposes.

Wittenbrink *et al.* (1996) claimed that the use of verity visualizations results in an integrated graphic 'so that users cannot help but interpret the resulting image holistically'. In this way, these graphics can be seen as examples of object or configural displays, which have been proposed for the integrated display of system information (Bennett *et al.* 1993, Carswell and Wickens 1996). These displays tend to have so called emergent features, such as shapes or patterns, that emerge through the interaction of elements of the graphical form, to provide easily perceptible information about system state or variable values. A 'star' display, in which system state variables graphed on different axes is an example. When state values are within normal parameters, a regular polygon shape is formed; deviations from normal are easily recognized. Research on object displays has concluded that merely presenting information in an integrated manner does not necessarily result in improved performance using the graphic, however. Performance improvements depend on the nature of the task and the degree to which task relevant features emerge from the graphic (Carswell and Wickens 1996).

Thus, while the displays of Wittenbrink *et al.* (1996) may provide potential benefits, it is necessary to consider and evaluate the degree to which such features are actually apparent to the users and the impact such display formats (as well as other formats for displaying uncertainty inherent in data) have on task performance. However, while there have been some limited subjective evaluations of these formats (e.g. asking the opinions of expert users; Wittenbrink *et al.* 1996), the differential impact of these representations of uncertainty on decision-making has not been assessed.

To date, then, there has been limited research on the effect of graphical representations of uncertainty and particularly representations of uncertainty regarding the identify of an object, on decision-making. An interesting question is how uncertainty regarding an objects' identify may be represented. One candidate representation is the use of blurred or degraded graphical images. MacEachren, (1992) suggested using blurred lines to indicate uncertain boundaries and blurred images or fill to indicate uncertainty in a type of land cover. Less resolution or geographic detail could be used to represent less certain data. Such a technique may be useful in conveying uncertainty because the manipulation of the image (e.g. through blurring or fuzziness) corresponds in a natural way to the level of uncertainty regarding the data. Norman (1988) discusses the advantage of natural mappings in the design of displays.

There is also some limited experimental evidence suggests that blurred or degraded graphical forms may be useful in conveying uncertainty. For instance, Lind *et al.* (1995) provided evidence that the form of displayed information may affect the use of uncertain data. In a study to investigate the extent to which the graphic depiction of weather systems could be degraded (due to technical limitations) and still be acceptable to general aviation pilots, Lind *et al.* found that pilots' estimates of weather hazards increased as the graphical distortion increased. In this case, the distortion took the form of larger polygon/ellipse shaped depictions of weather patterns, in contrast to the non-distorted continuous, fine-grained representation. This increase in perceived risk might indicate a decrease in participants' confidence of their understanding of the current weather patterns, perhaps indicating that the more degraded images might be useful in conveying the uncertainty regarding the weather patterns. Additionally, based on experimental evidence that a more concrete, graphical presentation of probabilistic information had a negative impact on decision-making performance compared to an alphanumeric presentation while a similar presentation of deterministic information had a positive impact, Kirlik (1995) hypothesized that the concrete display led participants to treat the probabilistic information as more certain than it really was. This observation is consistent with that of Kirschenbaum and Arruda (1994) who noted a non-significant trend for participants using a graphical representation to exhibit worse performance when the information regarding an objects' location was drawn from a less appropriate algorithm for computing position from raw sensor data. They suggested that this may have occurred because the elliptical presentation of uncertainty in target location could be more compelling and therefore misleading, than a verbal presentation, when the model used to generate the uncertain estimate was inaccurate. Such conclusions are consistent with recommendations from research in subjective probability and linguistic representations of probabilities, which have suggested that the presentation of uncertainty should be 'as precise as warranted by the available information' (Wallsten 1990: 28).

1.1. *Summary and proposed work*

In summary, extensive research in the areas of subjective probability and linguistic representations of uncertainty have confirmed the need to understand the impact of representations of uncertainty on decision-making. While results from research on subjective probability has suggested that the form of information representation is important for some tasks, research on linguistic representations of probability has indicated that while there may be individual differences in the interpretation of linguistic probabilities, performance on gambling-type tasks was relatively consistent across verbal and numeric representations of probabilities. Research on visualization of uncertainty in large data sets has focused primarily on visualization techniques rather than the impact of such techniques on decision-making. Finally, while some research on graphical representations uncertainty has suggested that such representations may provide value to decision-makers particularly under conditions of higher uncertainty, this research has had a limited focus on representation of uncertainty in object location and heading. There have been some indications that the use of degraded or blurred graphical forms may be an additional, natural way to represent uncertainty, which may be appropriate for representing uncertainty about the identity or type of an object. However, more research must be done to understand the effects such representations would have on decision-making.

Questions similar to those posed in the study of linguistic representations of uncertainty are appropriate in the study of graphical representations of uncertainty. Will decisions be equivalent across graphical and numeric representations of uncertainty, similar to the general finding regarding linguistic and numeric representations? How will individuals interpret uncertainty represented by different graphical forms, which could include degraded or blurred images? Will context have an effect on these interpretations? Similar to linguistic probabilities, will there be large individual differences in the manner in which graphical representations of uncertainty are interpreted? For the last question in particular, it may be that some kids of graphical representation have an advantage over linguistic representations of uncertainty. While graphical representations may not be as 'precise' as numeric representations (i.e. they may be interpreted as representing a range of probabilities), they may not be as vague as linguistic representations. People base their interpretation of linguistic probabilities on their experience with the meaning of the linguistic expressions. In many cases, there is no normative or definition based value corresponding to the linguistic phrases used. In contrast, graphical representations can be designed such that a normative assignment exists. For instance, the extent of blurring or degradation in a graphical form can be controlled by the designer to represent different levels of uncertainty.

To investigate these questions, a two-part study was performed which compared the use of numeric formats to degraded and blended icons for conveying situational estimates. For the first part, a classification study, five sets of thirteen more-or-less degraded and blended icons intended to represent levels of uncertainty regarding an object's identity were developed. Participants were asked to sort, order and rate icons in each set to investigate whether or not the icons could be distinguished and had the potential to represent different levels of uncertainty. Because of the fact that the icons were designed by changing the degree of degradation to correspond to the degree of uncertainty, it was hypothesized that participants would be able to order and rate the icons successfully.

A second experiment assessed the use of these icons in a decision-making task, in which participants had to rely on information conveyed by the icons to identify objects. The intent of this second experiment was to compare the impacts of graphical and numeric representations of uncertainty. If the use of blurred or degraded graphical formats conveys uncertainty in a natural and interpretable manner, then performance using the icons compared to numeric probabilities should be similar. If the level of degradation of the icons conveys the uncertainty or vagueness of the information more directly than precise, numeric values, then participants may behave as if they interpret the information as more uncertain.

The remained of the paper describes the two experiments.

2. Classification study method

For the classification study, participants were asked to sort, order and rate more or less degraded icons developed to indicate different levels of uncertainty.

2.1. *Participants*

Twenty participants, all undergraduate students, were paid $6.00 per hour for their participation in the classification study. The average age of participants was 20.7 years, with a range from 18 to 31 years. There were five female and 15 male participants. Participants came from various academic backgrounds, including engineering, medical fields and social sciences.

2.1.1. *Experimental design.* Five sets of icons were developed to represent the identity of an object as either hostile or friendly. Owing to prior work suggesting effects of task context on probabilistic decision-making (e.g. Brun and Teigen 1988, Dusenbury and Fennema 1996) different icon sets were chosen to see if any effects held across the particular image used and whether or not that image had prior cultural meaning for the participants. The icon sets were based on culturally understood symbols, abstract symbols or colour coded. Figure 1 shows the icons used in the experiment.

In order to represent the probabilistic nature of the information graphically, a series of thirteen icons were created to represent a range of probabilities (i.e. from $p(\text{Hostile}) = 0.0$ to $p(\text{Hostile}) = 1.0$). The iconic and abstract picture pairs were distorted and blended using a pixelizing function found in Adobe PhotoShop 4.0. For example, the 50% friendly/50% hostile picture blended both of the pictures in a pair

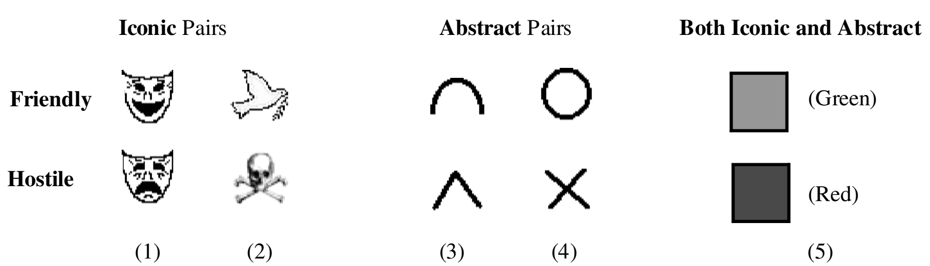

Figure 1. Five pairs of icons representing object identities as either hostile or friendly.

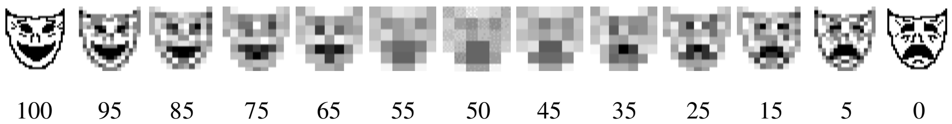

Figure 2. One series of 13 icons representing a range of probabilities that an object is hostile or friendly. The numbers indicate the probability that the object is friendly.

together. For the coloured icons, the series of icons was created by colouring each pixel in the icon either green or red based upon the probability desired. To illustrate how the pixelizing function works, the series of the distorted and blended icons for icon pair (1) are shown in figure 2.

2.1.2. *Experimental tasks.* Each participant performed a series of three tasks involving all five sets of icons. In addition to testing different icon sets and again due to prior research suggesting context effects in the interpretation of probability information, two task relevant contexts were compared. Ten participants performed the tasks under a 'friendly' framing context condition and ten participants performed the tasks under a 'hostile' framing context condition. In the friendly framing condition, participants were given task instructions that described the icons as more or less friendly. In the hostile framing condition, icons were described as more or less hostile.

The three experimental tasks were designed to measure whether the icons could be correctly sorted and assigned a probability rating according to the expected probabilities that the icons represented. Participants performed each of the tasks five times: once for each icon pair. The order of icon pairs was randomized across participants. The three experimental tasks were performed in order, as follows.

2.1.2.1. Sorting task. In the first task, a timed sorting task, participants were asked to sort cards into piles based on the icon shown on the card. The intent of this task was to insure that participants could distinguish among the icons in a set. If the icons could not be distinguished, then it would be impossible to conclude that they could be used to represent different levels of uncertainty. There were five instances each of the 13 possible icons in a set, for 65 cards. There were 13 reference cards, each showing one icon, taped to a desktop in a random order; participants were asked to create piles with cards that matched the reference card. The time to sort the cards and number of sorting errors was collected.

2.1.2.2. Ordering task. In the second task, participants were asked to order the set of thirteen pictures from most to least friendly (or hostile), depending on the framing condition. The intent of this task was to measure whether or not participants would mimic the order of the icons as intended when created. For the icon sets to be useful in conveying different levels of uncertainty, it was necessary for participants to order them appropriately. Participants were not told which icons corresponded to the hostile or friendly ends of the scale (e.g. they were not told that a circle represented a most friendly and an 'x', least friendly). Participants performed this task using a Visual Basic computer program, through which they could drag and drop the icons into the desired order. The ordering of the icons was recorded automatically by the computer.

2.1.2.3. Rating task. For the third task, participants were asked to rate each icon on continuous scale, with end-points of least and most friendly (or hostile). Participants marked their rating along a line connecting the endpoints; this distance was later measured and scaled based on the length of the line and used to identify their rating. The line was marked in increments of 10, from 0 (least) to 100 (most). As with the ordering task above, the intent of this task was to measure the extent to which participants' interpreted the meaning of the icons in a way that was consistent with expectations regarding the level of uncertainty that was being indicated with the icon. If participants' interpreted the meaning of the icons in ways that were very different from expectations it would make their use as representations of uncertainty suspect.

3. Classification study results

3.1. *Sorting task*

Mean time to sort cards based on the icon printed on the card ranged from 313.4 s (Circle/X pair) to 363.75 s (modified Military pair); however, the mean times did not differ significantly ($F_{4,95} = 0.7$; $p = 0.54$). A count of errors (e.g. number of missorted cards) showed 42, 49, 76, 31 and 36 errors for the masks, dove/skull, military, circle/x and colour pairs, respectively, indicating good sorting performance (an average of 48.6 errors out of 1300 cards sorted, or 3.6% errors). There were some differences in errors across icon sets. A χ^2-test indicated that the null hypothesis of equal error frequency across icon sets should be rejected ($\chi^2_4 = 26.64$, $p \leqslant 0$). In particular, more errors occurred for the military set. Participants were also asked to rate the difficulty of the sorting task by placing a mark on an 11 point scale with endpoints 0 (easy) and 10 (difficult). Average ratings were taken by measuring the placement of the mark with respect to the easy endpoint and scaling the measurement from 1 to 100. The average difficult rating was 40.0; there were no significant difference in difficult rating among icon sets ($F_{4,95} = 0.6666$; $p = 0.6167$). Thus, based on these dependent measures, the relative difficulty of identifying and sorting the thirteen icons did not appear to differ greatly across sets, although more errors were made with the modified military sets.

3.2. *Ordering task*

The order of the thirteen icons in each icon pair set was determined for each participant, for the hostile and friendly framing conditions, resulting in ten orders per icon pair (one from each participant) for each framing condition. Spearman correlation coefficients were computed comparing each participant's order to the expected order (based on the way the icons were created), for both framing conditions. As seen in table 1, absolute values of correlations for the hostile condition ranged from 0.005 to 1 and from 0.962 to 1 for the friendly framing condition. The correlations were similar across icon sets: average squared correlations across participants for the friendly framing condition were 0.98, 0.99, 0.92, 0.91 and 0.99 and were 0.46, 0.58, 0.54, 0.53 and 0.46 for the hostile framing condition, for the mask, dove/skull, military, circle/x and colour icon sets, respectively. There were two insignificant correlations, based on a significance level of $p = 0.01$, in the friendly framing condition and 22 in the hostile framing condition. Thus, ordering was more consistent and correct in the friendly framing condition than the hostile framing condition.

Negative correlations (seven in the friendly condition, six of which were significant and three in the hostile, none of which were significant) indicated that the

Table 1. Correlations between intended and actual orders by participants, across icon sets and framing conditions.

	Mask	Dove/skull	Military style	Circle/X	Colour
Friendly framing					
1	0.995	1	−1	1	0.995
2	0.989	0.995	1	1	1
3	1	1	0.995	1	1
4	1	1	−1	1	0.995
5	1	1	1	1	1
6	1	−1	−1	−1	−1
7	0.984	1	1	1	0.995
8	0.962	1	1	1	1
9	0.995	0.995	1	1	1
10	0.978	1	−0.44**	0.374**	1
Hostile framing					
1	0.126**	0.115**	0.115**	0.115**	0.115**
2	1	1	1	1	0.995
3	0.566	−0.038**	0.544	−0.297**	0.665
4	0.412**	0.093**	0.115**	0.148**	0.088**
5	0.005**	0.714	0.099**	0.044**	0.181**
6	0.978	1	1	1	0.434**
7	0.148**	1	1	0.995	0.456**
8	0.978	1	0.978	0.995	0.989
9	0.995	0.995	1	1	1
10	−0.165**	0.516	0.28**	0.44**	0.835

** Insignificant correlations.

participant reversed the hostile and friendly ends of the scale (they were not told which icons corresponded to which endpoints before the experiment). It is interesting to note that even for the two ‘abstract’ icons, reversals happened at a rate less than chance. When developing the icon sets, we anticipated that participants would associate typical cultural meanings to the dove/skull, mask and colour (red/green) sets, although it seems highly unlikely that they would have used such icons to classify objects on a radar screen prior to this experiment. However, we did not expect there to be an obvious meaning to the circle/X and military style sets, given the abstract nature of the circle and X graphics and the fact that the participant pool was a university rather than military setting. The results indicate, however, that there was some meaning intrinsic to the more abstract icon pairs. Perhaps the X had a more negative connotation due to the use of similar graphics to indicate prohibited actions on things like signage, while the ‘sharp’ point of the chevron in the military style icon set indicated hostility. Additionally, it is possible that participants may have had exposure to some military-style icons through experiences like video games. Regardless of reason, it did appear that the abstract pairs were more meaningful than was initially intended.

As with the Sorting task, subjective ratings of difficulty were collected. Average difficulty was 28.9 out of 100; there were no significant differences in average difficulty between icon sets ($F_{4,95} = 1.196$; $p = 0.3177$).

3.3. *Rating task*

Participants’ ratings of each icon in an icon set along a most to least friendly (or hostile) continuum were correlated with the expected ratings based on icon design

Table 2. Correlations between intended probabilities and participant ratings across icon sets and framing conditions.**

	Mask	Dove/Skull	Military style	Circle/X	Colour
Friendly framing					
1	0.952	0.901	−0.67	0.044**	0.113**
2	0.993	0.953	0.984	0.966	0.995
3	0.908	1	0.964	0.967	0.977
4	0.995	0.966	0.999	0.995	0.989
5	0.993	0.99	0.974	0.964	0.919
6	0.444**	−0.998	−0.967	−0.984	0.959
7	0.996	0.989	0.971	0.956	0.962
8	0.956	0.96	0.953	0.907	0.938
9	0.843	0.938	−0.858	0.964	0.714
10	0.94	0.956	−0.893	0.786	0.949
Hostile framing					
1	−0.11**	0.061**	−0.11**	−0.033**	−0.061**
2	0.978	0.945	0.973	0.995	0.962
3	0.339**	0.383**	−0.059**	0.248**	0.864
4	0.195**	0.18**	0.242**	0.256**	0.113**
5	0.247**	0.427**	0.154**	0.257**	0.308**
6	0.961	0.979	0.951	0.89	0.834
7	0.506	0.966	0.603	0.839	0.113**
8	0.275**	0.176**	0.214**	0.41**	0.889
9	0.993	0.981	0.992	0.995	0.988
10	−0.371**	0.49**	0.264**	0.485	0.801

** Insignificant correlations.

(e.g. representing 0, 5, 15%, etc.). Spearman correlation coefficients are shown in table 2, for both framing conditions. As with the ordering task, correlations tended to be high and significant in the friendly framing condition and more variable in the hostile framing condition. There were three non-significant correlations in the friendly framing condition and 27 non-significant correlations in the hostile framing condition. Thus, participants in the friendly framing condition appeared to rate icons in a manner consistent with prior expectations. The presence of high negative correlations in the friendly framing condition (there were six significant negative correlations) again indicates that some participants reversed the meaning of the endpoints.

For each participant, the maximum and minimum rating values assigned within each icon set were identified, to see if participants would rate the icons across the entire continuum. Average maximum and minimum values for each icon set, by framing condition, are displayed in figure 3 and show that the range of ratings for friendly framing condition was greater across all icon pairs, due to the tendency to make less extreme ratings, particularly of high hostility, in the hostile rating condition.

3.4. *Classification study discussion and conclusions*

The goals of the initial classifications study were to determine if participants could distinguish among the degraded icons, order them as intended and to match them to the level of uncertainty that was intended and to see if performance on these factors differed across the icon sets used. Without such an initial study, it would be difficult to determine whether any performance differences using the icons compared to other

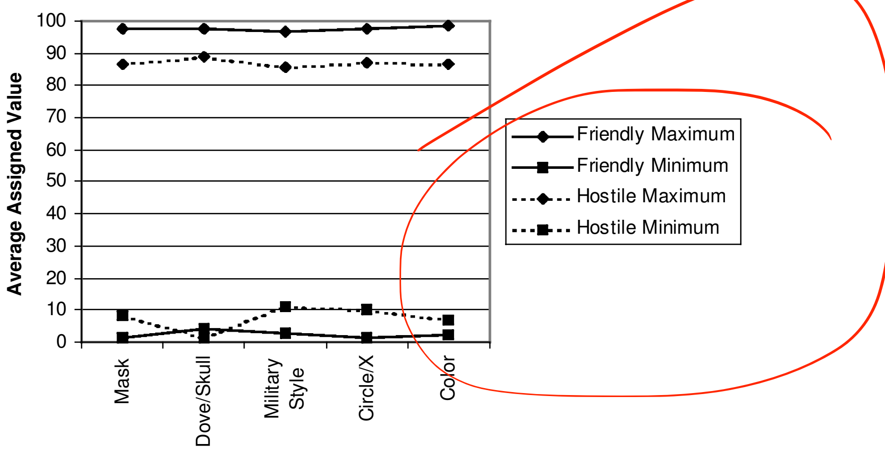

Figure 3. Average maximum and minimum ratings given for each icon category, for the Friendly and Hostile framing conditions.

(e.g. numeric) representations in a decision-making context were due to the manner in which participants interpreted the level of uncertainty, or due to the fact that they had difficulty simply distinguishing icons, or identifying which icons represented objects with more uncertainty than others. Based on the results of the classification study, we concluded that participants could sort, order and rate the icon sets as intended in the friendly framing condition. Additionally, there were few differences among icon sets. However, there appeared to be performance differences between the hostile and friendly framing conditions for the ordering and rating tasks. One possible explanation for this is that the difference is reflecting the between-subjects nature of the framing condition: perhaps participants systematically differed in their abilities to perform the task, between task conditions. From tables 1 and 2, it appears that five participants (1, 3, 4, 5, 10) in the negatively framed condition consistently had difficulties both ordering icons and rating icons, across icon sets. However, since participants were randomly assigned to framing conditions, there is no reason to expect a systematic difference. Alternatively, it could be that certain participants were less comfortable ordering or rating icons in terms of hostility—perhaps, as indicated by the range of ratings in the rating task, they did not want to assign as extreme ratings and in the both the ordering and rating task, did not want to indicate that some icons were more hostile than others and thus responded more randomly in that task. Further research using framing as a within-subjects condition could shed additional light on these findings. However, based on these differences, a friendly framing was chosen for the subsequent display study. Owing to the similarity in performance across icon sets, there was no particular reason to select one set over another; the dove/skull, modified military and colour icon sets were selected for further study.

One potential limitation with the classification study is that participants performed the tasks in order. However, since participants were exposed to many different icons in each task, it seems unlikely that memory for or experience with a particular icon carried over and influenced performance from task to task. In particular, participants were shown the icons in the rating task in random order, one at a time, reducing the chance that their memory for the order they had assigned influenced their ratings; a lack of improvement across the three tasks lends support for the conclusion that experience with the icon sets from task to task was not a strong contributor to task performance.

4. Decision-making study method

Results from the classification study indicated that sets of distorted icons could be appropriately ordered and span a range of descriptive levels. A second study was conducted to determine the potential utility of these icons on a dynamic decision-making task.

4.1. *Participants*

Ninety-four volunteers were recruited from the university population for this study and were paid $6.50 per hour for their participation. Average participant age was 26.7 years, with a range from 18 to 41 years. There were 31 female and 65 male participants. Participants came from a variety of academic backgrounds including engineering, business, medical fields, social sciences and the arts.

4.2. *Experimental design*

Participants in the experiment were asked to identify objects as hostile or friendly in a dynamic, graphical environment (figure 4). The only information participants had about the objects was a probabilistic estimate of its identity. These estimates changed over the course of the trial.

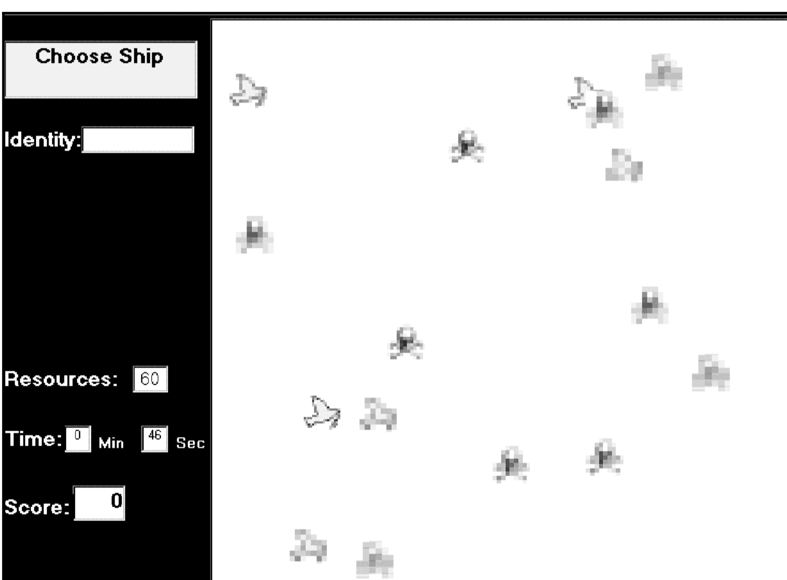

Figure 4. Interface for the decision-making experiment. Entities represented by icons in the right portion of the screen move in random patterns. To identify an object, participants clicked on an object to select it and then entered an identification on the left portion of the screen.

There were two primary manipulations tested in this study: picture type and display type. Three picture pairs were chosen from those used in the classification study: the dove-skull, modified military (Abstract pair #3 in figure 1) and red-green colour pair. Four display manipulations to present the probabilistic estimates about objects' identities were also investigated, as follows.

- Degraded image only, no probability instruction condition: In this condition, an icon representing the object became more or less degraded to indicate the current probabilistic estimate that the object was friendly. No numeric probabilities were shown on the screen. Task instructions in this condition did not show probabilities associated with the icons.
- Degraded image only, with probability instruction condition: In this condition, like the first, only degraded icons were seen during the task; however, participants were shown the probabilities together with the icons in the task instructions.
- Degraded image with probability condition: In this condition, in addition to the icons changing in degradation level, icons were annotated with a numeric probability corresponding to the current probability the object was friendly. Task instructions were similar to the second condition.
- Probability only condition: Finally, in this fourth condition, objects were indicated by a static, non-degraded icon and a numeric probability that changed throughout the trial. The static, non-degraded icon used was either the hostile or friendly endpoint icon, depending on the current estimate (i.e. for estimates $< 50\%$ friendly, the hostile endpoint icon was used, otherwise, the friendly icon was used). Thus, in this condition, the only indication of the level of uncertainty was the numeric probability. The form of this condition was chosen rather than just presenting a probability by itself, to keep the general form of the displayed information as similar as possible to the other task conditions.

Example icons are shown in figure 5. These manipulations were intended to investigate if degraded images could be used as a viable alternative to numeric

Display Type	Example Stimulus
Degraded Only*	
Degraded with Probability	
Probability Only	

Figure 5. Example display stimuli for the degraded only, degraded with probability, and probability only display conditions.

probabilities on displays and if the use of such images, either alone or in combination with numeric probabilities, could cause decision-makers to behave differently (e.g. more conservatively with respect to the uncertain information) than if only numeric probabilities were available.

The icon type and display type conditions were crossed for 12 between-subject factors. There were eight participants per condition. Each participant completed four trials.

4.3. *Experimental task*

Participants were asked to identify the status of objects as either hostile or friendly. There were 40 objects per trial, for four trials. Each object was randomly assigned an initial probability of being either friendly or hostile and also an actual identity, displayed with the appropriate icon. All 40 objects were displayed from the start of the trial and had a randomly assigned probability of being friendly. Throughout the trial, this initial probability periodically changed (at random intervals between 30 and 90 s) to become more indicative of the icons' actual identity (e.g. move more toward either 0 or 100% friendly) 70% or the time and less indicative 30% or the time. Thus, with time, identities became increasingly certain; however, any one change was not completely predictive of the object's actual identify. Estimates of objects' identities could range from a 5% chance of being friendly, to a 95% chance of being friendly—the estimates did not reach 0 or 100%. Objects moved in a random pattern around the screen.

Each trial had a maximum time limit of ten minutes. Participants were assigned 60 resources at the beginning of each trial. Participants lost one resource for every correct identification and two resources for every incorrect identification. Thus, at least one resource was required for each identification. Additionally, remaining resources corresponding to 16% of the unidentified contacts were lost each minute. Overall score per trial was computed as the number of correct identifications plus the resources remaining at the end of the trial, multiplied by a factor of 10. The intent of these manipulations was to reward participants' waiting until they were confident of an object's identity (so they would not make an incorrect identification) but to not wait longer than necessary (or resources would be lost). Losing resources not only lowered the overall score, but could make it impossible to identify any more objects, since each identification required at least one resource. This scoring scheme was intended to duplicate aspects of real world decision-making, in which information can become more certain over time (as more information is gathered), but where there are time constraints necessitating action before all aspects of a situation are known with certainty.

4.4. *Procedure*

Participants were given task instructions to read as the experimenter read them aloud. These materials included, as noted above, pictures of the icon sets, labelled with the associated probabilities for the second and third condition and labelled with endpoints of 'Friendly' and 'No Chance Friendly' for the first condition. For the fourth condition, participants were shown a set of non-degraded endpoint icons, labelled with the range of probabilities. Participants could refer to the task instructions during the experiment. The experimenter also assisted participants in performing a short training trial, similar to the experimental trial. Participants then performed the four trials.

Data, including the time for each trial, participants' scores and point at which each object was identified, were automatically captured as the trials were run.

5. Decision-making study results

Several dependent measures were computed and analysed to understand participants' overall task performance, as well as aspects of their decision-making behaviour.

5.1. *Performance measures*

Trial score, trial time, number of identifications and percent correct were analysed using a three-factor, repeated measures ANOVA. Icon type (three levels) and Display Format (four levels) were treated as between-subjects factors, while trial (four levels) was treated as a within-subjects factor.

5.1.1. *Trial score.* Trial score, which combined resource use with identification performance, was computed as described in the previous section. There was a significant effect of display type ($F_{3,84} = 3.192$, $p = 0.028$) and trial ($F_{3,252} = 47.438$, $p < 0.000$) on trial score. As shown in figure 6, trial score tended to increase over the first three trials, while trial score was highest for the degraded icon only-no probability directions condition and similar for the other three conditions. There was no significant effect of icon type ($F_{2,84} = 0.406$, $p = 0.667$) and no interactions were significant.

5.1.2. *Trial time.* The total time for each trial was significantly affected by display type ($F_{3,84} = 9.537$, $p < 0.000$) and trial ($F_{3,252} = 14.848$, $p < 0.000$). As shown in figure 7, trial time tended to be lower for the degraded icon only-no probability directions condition and similar for the other three conditions. Trial time also tended to decrease over trials. There was no significant effect of icon type ($F_{2,84} = 1.81$, $p = 0.312$) and no interactions were significant.

5.1.3. *Number of identifications.* The total number of identifications, whether correct or incorrect, made in each trial was also investigated. Recall that there were potentially 40 objects per trial which could be identified; however, participants' may have identified fewer because they ran out of time, or did not feel confident

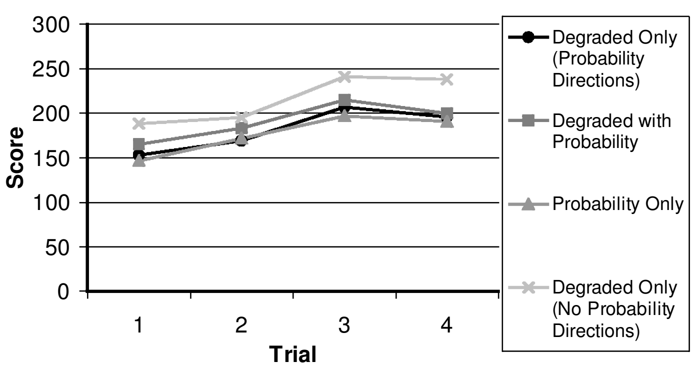

Figure 6. Overall score per trial, by display condition.

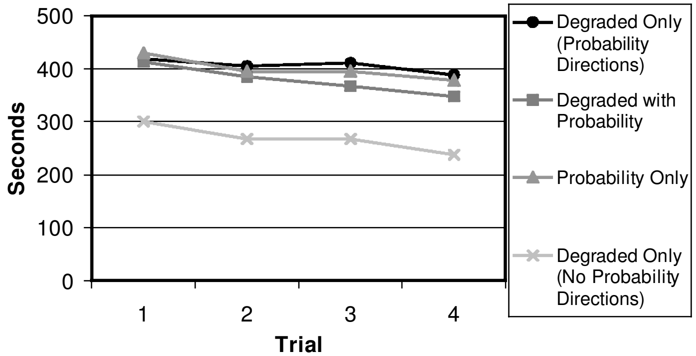

Figure 7. Time per trial, by display condition.

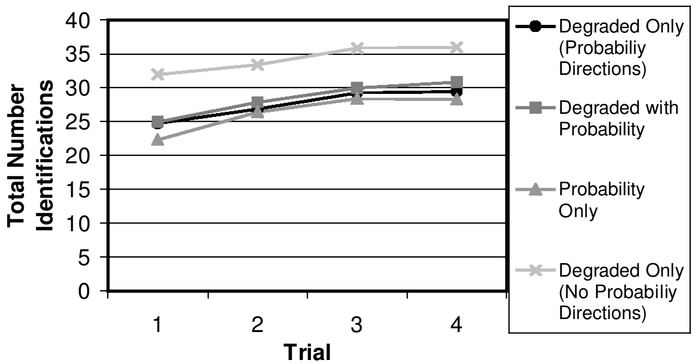

Figure 8. Total number of identifications per trial, by display condition.

enough in their knowledge about an object's identity to make an identification. Again, there was a significant effect of display type ($F_{3,84} = 6.991$, $p < 0.000$) and trial ($F_{3,252} = 48.768$, $p < 0.000$) on trial score. As shown in figure 8, the total number of identifications tended to increase across trials, while the number of identifications was highest for the degraded icon only-no probability directions condition and similar for the other three conditions. There was no significant effect of icon type ($F_{2,84} = 0.786$, $p = 0.459$) and no interactions were significant.

5.1.4. *Per cent correct.* The percent of objects correctly identified (computed as the number of correct identifications divided by the number of identifications made in a trial) was significantly affected by trial ($F_{3,252} = 9.185$, $p < 0.000$). There was no significant affect of display type ($F_{3,84} = 2.036$, $p = 0.115$) or icon type ($F_{2,84} = 0.033$, $p = 0.967$); no interactions were significant. As seen in figure 9, the percent correctly identified tended to increase from the first and second trial to the third and fourth trial. Also, though the differences were not significant, there seemed to be a trend for better performance for the combined degraded image-probability display type and somewhat lower percent correct for the degraded icon only-no probability directions condition.

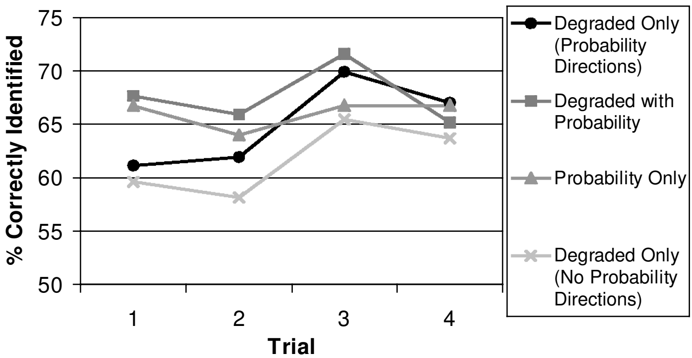

Figure 9. Per cent correct identifications per trial, by display condition.

5.2. *Strategy measures*

In addition to the performance measures described above, participants' actions were analysed to determine at what level of uncertainty they chose to identify icons. In particular, the number of times participants made an identification when an object was represented by each of the different icons or probability levels was examined. The intent of these measures was to determine if the icon type or display type factors influenced decision behaviour in terms of the level uncertainty in the situation assessment that was 'acceptable' to participants with respect to making an identification decision. This was analysed for objects identified as hostile, friendly and across all identifications.

Contingency table analyses were performed to investigate whether the proportion of objects identified at each probability level differed according to icon type, or display type, for friendly identifications, hostile identifications and all identifications. While, analyses showed a significant relationship between icon type and probability

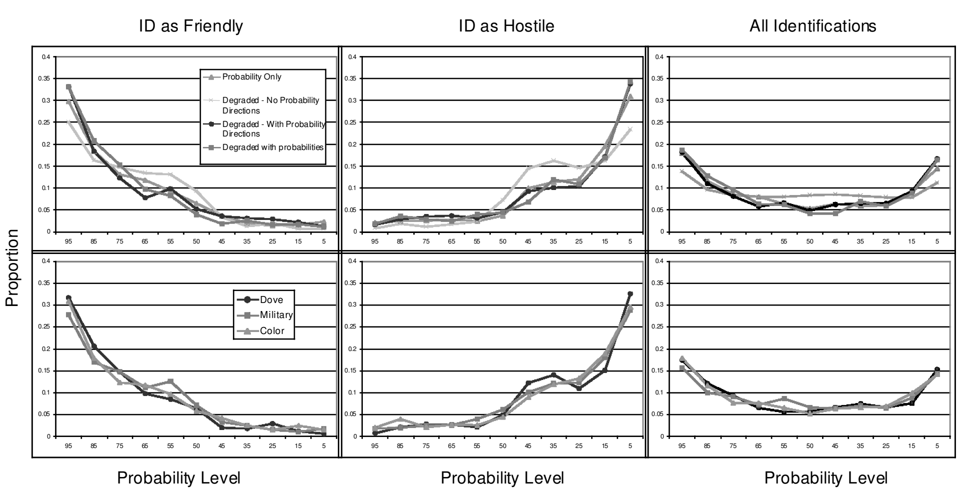

Figure 10. Scaled count of identifications at each level of uncertainty, for friendly identifications, hostile identifications and all identifications.

level ($\chi^2_{20} = 101.5$, 71.4, 71.46; $p < 0.001$, for friendly, hostile and all identifications, respectively), inspection of figure 10 shows few consistent differences in the proportion of identifications made at each uncertainty level across icon types, for friendly, hostile and all identifications. Analyses of display type crossed with probability level also showed significant effects ($\chi^2_{30} = 186.67$, 189.72, 237.19, $p < 0.001$, for friendly, hostile and all identifications, respectively). Inspection of figure 10 shows some patterns that may explain this result. For instance, there was a trend for participants in the degraded icon only-no probability directions condition to identify a greater proportion of objects at lower levels of certainty (e.g. icons closer to 50%) than participants in the other conditions, while participants in the degraded icons with numeric probabilities tended to identify objects at points of greater certainty (e.g. closer to 5 or 95%).

6. Discussion

6.1. *Classification study*

The goals of the classification study were to investigate whether or not sets of degraded icons could be used to convey situational estimates; specifically, estimates of whether an object was particular type (hostile or friendly). Results from the sorting task showed that there were few (an average of 3.6%) errors sorting the cards, indicating that the icons in the sets could be successfully distinguished, ordered and assigned situational ratings. Performance on the ordering task showed that in the friendly framing condition, participants could arrange icons according to their intended order. Finally, results from the rating task indicated high and significant, correlations between ratings given by participants and the intended estimates in the friendly condition. Thus, overall, it appears that in at least one context, icons could be successfully distinguished, ordered and assigned situational ratings.

Different types of icons were tested to in the classification study to determine the presence or absence of meaning associate with the icons would affect results. Additionally, different icon sets were tested to investigate the potential generalizability of results from the classification study and subsequent decision-making study. Specifically, we investigated whether study results appeared to be affected by the choice of icon type, or icon set, used. Results from the three classification study tasks showed few differences in performance across icon sets: sorting time, subjective difficulty ratings, order and rating correlations and rating midpoints and spans were similar across all icon sets. One difference was the number of sorting errors: it appeared to be more difficult to sort the military-style icons, which may be due to the similarity of the two end-point icons. Thus, results suggest that the concept of using degraded and blended icons to convey situational estimates may be useful across different types and instances of icons.

While there were few differences in performance due to icon sets, there were differences between the hostile and friendly framing contexts. As noted above, individual participants in the hostile framing context appeared to have difficulty both ordering icons and rating icons, across icon sets. Participants' ordering and rating of icons were less likely to be similar to the expected ordering in the hostile condition. Additionally, participants' ratings of icons spanned a smaller range in the hostile framing condition than the friendly framing condition, perhaps reflecting a hesitancy to make more extreme assessments of high or low hostility. Interestingly, this is similar to results described by Jian *et al.* (2000) who found that people's ratings of the similarity between test words and the concept of distrust, were less extreme than ratings of the same test words and concepts of trust, a more positive concept. In both cases, participants were less extreme in making assessments of the more negative concept, perhaps reflecting a more general tendency to avoid strong statements regarding less pleasant situations. Banbury *et al.* (1998) also showed a friendly-hostile framing effect: participants were less likely to shoot at a target when probability it was friendly was explicitly stated in addition to the probability that it was hostile. These results underscore the importance of recognizing how context can impact decision-making and the interpretation of information, as previously noted in studies of judgement and linguistic probability representations (Brun and Teigen 1988, Dusenbury and Fennema 1996). Future work using context as a within-subject condition could further clarify the effects of this particular set of contexts.

6.2. *Decision-making study*

The goals of the display study were to test the effect of different graphical and numeric representations of situational uncertainty on a decision-making task. Similar to the classification study, results across dependent measures indicated little or no effect of icon set on performance, supporting the conclusion that task performance was not dependent on the particular icon sets tested.

There were, however, some significant effects of display type. Participants who saw only degraded icons as situational estimates, with no numeric probabilities on the screen or in the task instructions, performed differently than participants who saw both degraded icons and probabilities, non-degraded icons with changing probabilities, or degraded icons with the probability-icon mapping illustrated in the task instructions. Specifically, participants had a higher overall score than participants in other conditions, although the percent of identifications made correctly did not differ significantly. These results are consistent with and can be explained by, patterns of results from the other dependent measures. Participants in the degraded icon only-no probability directions condition had a significantly better score, even though they did not identify more objects correctly, because they identified more icons per trial and finished trials sooner than participants in other conditions, thus conserving more resources at trial end (a part of the score). Examination of figure 10 suggests a trend in terms of when participants made identifications that is also consistent with these results. In the degraded icon only-no probability directions condition, participants appeared to make more identifications when objects had more uncertain situational estimates (that is, when icons or numeric probabilities were closer to 50%). Since icons representing objects tended to become more certain as the trial progressed, this indicates that participants were making their identification decisions sooner, thus resulting in shorter trial times and a greater number of identifications per trial.

Overall, these results suggest that different displays of the uncertainty regarding objects' identity did impact participant's decisions about object identifications. Participants who only had information presented with the degraded icons appeared to make decisions with less certain information, resulting in significantly faster trial times and overall trial scores. That is, there was a trend for these participants to be less conservative than participants who had access to the numeric probabilities, although this strategy did not significantly affect the percentage of correct identifications made. Thus, one can conclude that the use of degraded images to represent situational uncertainty can improve certain aspects of decision-making performance, while not significantly reducing performance on other aspects. In some ways, then these results are consistent with results from the literature on linguistic representations of uncertainty, which have shown few differences between linguistic and numeric representations of uncertainty on judgement tasks. Participants were able to perform as well with the graphical representation as with those that included numeric information. Researchers in linguistic probabilities have suggested that to use vague representations of uncertainty (such as linguistic phrases) to reason in decision-making tasks, people 'sample' a point from a range of probabilities they believe to be represented by the phrase. The sampling range is based on the shape of the membership function relating the linguistic phrase to the probability values and is theorized to extend from a central value to the peak membership value (Wallsten and Budescu 1995). An analogous explanation in this context is that participants mapped the graphical representation to a range of probability values, but sampled probabilities reflected higher certainty than was normatively expected when deciding to make identifications, thus resulting in performance that was less conservative than participants who saw the numeric values. Further work that explicitly elicits membership functions for the graphical representations, analogous to those developed for linguistic phrases, could shed further light on these effects.

Additionally, performance of participants who saw only degraded icons on the screen, but who were given the icon-probability mapping in task instructions, was similar to performance of participants who saw both icons and probabilities on the screen, across all dependent measures. Thus similarity is likely due to the fact that participant could refer to the task guidelines during the task and thus treated the icons similarly to the icon with probability condition.

Further examination of figures 9 and 10 suggest additional, interesting trends. While figure 9 suggests there was a trend for participants in the degraded icon only-no probability directions condition to have a lower percentage of correct identifications, participants who saw both the degraded icons and numeric estimates on the screen tended to have better identification performance for three of the four trials. Additionally, it appears from figure 10 that participants who saw the combined representation, or who saw the degraded icons on the screen, but received the icon-probability mapping in the task instructions, tended to act most conservatively, making more identifications at the most certain icon and probability levels (5 and 95%). Thus, a representation that combines a graphical image along with a numeric probability may cause decision-makers to act more conservatively, waiting until information is more certain, thus improving decision-making performance.

6.3. *Implications and future work*

The results of these studies have important implications for the design of displays and for future research regarding the display of uncertain information. While previous studies have focused on using graphical formats to display uncertainty regarding the position or intended heading of an object, this study considered the use of a graphical format to display uncertainty associated with a diagnostic assessment of an object's identity. We attempted to convey the degree of uncertainty through the use of more or less degraded, or in the case of the colour set, blended icons. The studies found that such icons provided an advantage on some aspects of performance and relatively equivalent performance on other aspects. Additionally, participants who had access to the mapping between the icons and probabilities, but who only saw the degraded icons on the screen, had similar performance to participants who saw numeric probabilities. These results imply that that the use of such icons may provide a means of conveying uncertainty about a situation.

Given the positive outcome of this initial study, future research should be conducted to determine the suitability of such display techniques across different and more realistic task situations. For instance, such an approach may be particularly valuable for applications, such as defence applications, where there are many objects with associated uncertainty to be display and the addition of on-screen numeric probabilities would add to screen clutter. Use of degraded icons would still providing a constant reminder of the level of uncertainty associated with the information available about an object. Future research should investigate potential advantages of this display approach under conditions of display clutter.

Additionally, there was some evidence to suggest that the use of such icons in combination with numeric probabilities may cause decision-makers to adopt a more conservative approach, waiting for information to become more certain before making a decision. Thus, in cases where such conservatism is desirable, displays which augment numeric estimates with appropriately degraded graphical icons may be useful.

It is important to note that in the task studied here, multiple (up to 40) entities with corresponding icons were displayed on the screen simultaneously. Thus, participants were able to judge the level of uncertainty regarding any entities identity with respect to other entities on the screen. Thus, even in the degraded-icon only condition, they could judge the relative level of uncertainty of an entity and perhaps use icons that appeared very blurred, or very certain, to judge the absolute level of certainty of other icons (e.g. determine how much more blurred one icon was than the sharpest icon). This fact may have contributed to the ability of participants to successfully perform the task in the degraded icon only condition. Future studies should investigate whether similar effects hold in tasks where participants can only see one icon at a time and thus must make a judgement of uncertainty based solely on the level of degradation rather than a comparison to other icons.

Finally, an additional question of interest is the number of icons chosen to represent levels of uncertainty. Examination of figure 10 shows a pattern, particularly for hostile identifications, that the counts of identifications tended to peak at the most certain level (e.g. 5% chance of being friendly), level off from 15 to 45% chance of being friendly and then drop to a consistently low level for probabilities >0.5. This indicates that participants may have considered icons and probabilities to fall into three categories: very certainly hostile, somewhat certainly hostile and certainly not hostile. Additional studies should be conducted to compare decision-making performance when the scale used to measure and display uncertainty is more or less finely grained, both to understand the levels between which human decision-makers distinguish and to determine the detail with which uncertainty estimates must be computed and presented.

7. Conclusions

Results from the classification study indicated that sets of degraded or blended icons intended to represent levels of situational uncertainty could be ordered and rated in a manner similar to expectations. Additional results from the classification study indicated a framing affect on performance: participants' interpretations of displayed information became less ideal in a negatively framed context. Across both studies, the lack of differences in performance among icon types or forms provides evidence that the experimental results are not specific to a particular icon form.

Results from the subsequent decision-making study showed that participants using displays with only degraded icons performed better on some performance measures and as well on other measures, than participants who saw degraded icons annotated with numeric probabilities, where information regarding uncertainty was conveyed via numeric probabilities, or where numeric probabilities were mapped to the icons in the task instructions. That is, the presence of numeric probabilities did not provide a statistically significant advantage in this task. These results are significant, because they indicate both that people can understand uncertainty conveyed through such a manner and thus that the use of distorted or degraded images may be a viable alternative to convey situational uncertainty, particularly in circumstances where the addition of on-screen numeric probabilities would add to screen clutter and where a constant reminder of the level of uncertainty associated with the information available about an object is desired. Further examination of the results indicated that there were trends for participants who saw degraded icons in addition to numeric probabilities to behave more conservatively, waiting for more certain information before making decisions. Thus, degraded icons may also be useful in concert with numeric presentations of uncertainty.

Acknowledgements

The authors acknowledge the support of the Air Force Research Laboratory, Human Effectiveness Directorate (G. Kuperman, sponsor), the UB Center for Multi Source Information Fusion (J. Llinas, Director) and the National Science Foundation under Grant No. IIS9984079 for their support of the work, as well as the comments of anonymous reviewers on an earlier version of this paper.

References

ANDRE, A. D. and CUTLER, H. A. 1998, Displaying uncertainty in advanced navigation systems, *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (Santa Monica: Human Factors and Ergonomics Society), 31–35.

AYTON, P. and WRIGHT, G. 1994, Subjective probability: what should we believe? in G. Wright and P. Ayton (eds), *Subjective Probability* (Chichester: Wiley), 163–183.

BAR-HILLEL, M. 1980, The base-rate fallacy in probability judgements, *Acta Psychologica*, **44**, 211–233.

BENNETT, K. B., TOMS, M. L. and WOODS, D. D. 1993, Emergent features and configural elements: designing more effective configural displays, *Human–Computer Interaction*, **35**, 71–97.

BRUN, W. and TEIGEN, K. H. 1988, Verbal probabilities: ambiguous, context-dependent, or both?, *Organizational Behavior and Human Decision Processes*, **41**, 390–404.

BUDESCU, D. V. and WALLSTEN, T. S. 1985, Consistency in interpretation of probabilistic phrases, *Organizational Behavior and Human Decision Processes*, **36**, 391–405.

BUDESCU, D. V. and WALLSTEN, T. S. 1990, Dyadic decisions with numerical and verbal probabilities, *Organizational Behavior and Human Decision Processes*, **46**, 240–263.

BUDESCU, D. V. and WALLSTEN, T. S. 1995, Processing linguistic probabilities: general principles and empirical evidence, in J. Busemeyer, D. L. Medin and R. Hastie (eds), *Decision Making from a Cognitive Perspective* (San Diego: Academic Press), 275–318.

Budescu, D. V., Weinberg, S. and Wallsten, T. S. 1988, Decisions based on numerically and verbally expressed uncertainties, *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 281–294.

Butt, J. L. 1988, Frequency judgements in an auditing-related task, *Journal of Accounting Research*, **26**, 315–330.

Carswell, C. M. and Wickens, C. D. 1996, Mixing and matching lower-level codes for object displays: evidence for two sources of proximity compatibility, *Human Factors*, **38**, 1–22.

Christensen-Szalanski, J. J. J. and Beach, L. R. 1982, Experience and the base-rate fallacy, *Organizational Behavior and Human Performance*, **29**, 270–278.

Christensen-Szalanski, J. J. J. and Bushyhead, J. B. 1981, Physicians' use of probabilistic information in a real clinical setting, *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 928–935.

Cosmides, L. and Tooby, J. 1996, Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under uncertainty, *Cognition*, **58**, 1–73.

Dusenbury, R. and Fennema, M. G. 1996, Linguistic-numeric presentation mode effects on risky option preferences, *Organizational Behavior and Human Decision Processes*, **62**, 109–122.

Erev, I. and Cohen, B. L. 1990, Verbal versus numerical probabilities: efficiency, biases and the preference paradox, *Organizational Behavior and Human Decision Processes*, **45**, 1–18.

Gigerenzer, G. 1994, Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa), in G. Wright and P. Ayton (eds), *Subjective Probabilities* (Chichester: Wiley), 129–161.

Gigerenzer, G. and Hoffrage, U. 1995, How to improve Bayesian reasoning without instruction: frequency formats, *Psychological Review*, **102**, 684–704.

Jian, J. Y, Bisantz, A. M. and Drury, C. G. 2000, Foundations for an empirically determined scale of trust in automated systems, *International Journal of Cognitive Ergonomics*, **4**, 53–71.

Kirlik, A. 1995, The design of perceptually augmented displays to support interaction with dynamic systems, *Proceedings of the 6th IFAC Symposium on the Analysis, Design and Evaluation of Man–Machine Systems*, 809–814.

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

Koehler, J. J. 1996, The base rate fallacy reconsidered: descriptive, normative and methodological challenges, *Behavioral and Brain Sciences*, **19**, 1–53.

Lind, A. T., Dershowitz, A., Chandra, D. and Bussolari, S. R. 1995, The effect of data link-provided graphical weather images on pilot decision-making, in *IFAC Proceedings*—1995.

MacEachren, A. M. 1992, Visualizing uncertain information, *Cartographic Perspective*, **13**, 10–19.

Norman, D. A. 1988, *The Psychology of Everyday Things* (New York: Basic Books).

Pang, A. T., Wittenbrink, C. M. and Lodha, S. K. 1997, Approaches to uncertainty visualization, *Visual Computing*, **13**, 370–390.

Solomom, I., Ariyo, A. and Tomassini, L. A. 1985, Contextual effects on the calibration of probabilistic judgements, *Journal of Applied Psychology*, **70**, 528–532.

Tversky, A. and Kahneman, D. 1982a, Availability: a heuristic for judging frequency and probability, in D. Kahneman, P. Slovic and A. Tversky (eds), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press), 163–178.

Tversky, A. and Kahneman, D. 1982b, Evidential impact of base rates, in D. Kahneman, P. Slovic and A. Tversky (eds), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press), 153–162.

Tversky, A. and Kahneman, D. 1982c, Judgment under uncertainty: heuristics and biases, in D. Kahneman, P. Slovic and A. Tversky (eds), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press), 3–22.

Tversky, A. and Kahneman, D. 1982d, Judgments of and by representativeness, in D. Kahneman, P. Slovic and A. Tversky (eds), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press), 3–20.

WALLSTEN, T. 1990, The costs and benefits of vague information, in R. M. Hogarth (ed.), *Insights in Decision Making* (Chicago: University of Chicago Press), 28–43.

WALLSTEN, T. S. and BUDESCU, D. V. 1995, A review of human linguistic probability processing: General principles and empirical evidence, *Knowledge Engineering Review*, **10**, 43–62.

WALLSTEN, T. S., BUDESCU, D. V., RAPAPORT, A., ZWICK, R. and FORSYTH, B. H. 1986, Measuring the vague meanings of probability terms, *Journal of Experimental Psychology: General*, **115**, 348–365.

WITTENBRINK, C. M., SAXON, E., FURMAN, J. J., PANG, A. T. and LODHA, S. K. 1996, Glyphs for visualizing uncertainty in environmental vector fields, *IEEE Transactions on Visualization and Computer Graphics*, **2**, 266–279.

About the authors

Ann M. Bisantz is an assistant professor in the Department of Industrial Engineering at the University at Buffalo, State University of New York, where her research interests include human decision making and information displays. She received a PhD in Industrial and Systems Engineering from the Georgia Institute of Technology in 1997 in the area of human–machine systems.

Richard Finger received his MS degree in Industrial Engineering from the University at Buffalo in 2001, with a concentration in Human Factors. He is currently working in Human Factors for Accenture.