Experiments exploring interactions between social and decision-making factors in a maritime military task.

Charlotte E. R. Edmunds, Sarah Jenkins, Adam Harris, Magda

Osman

Queen Mary, UCL, University of London

In the first set of experiments, we are in the process of investigating whether and how presenting uncertainty affects participants’ performance in a change detection task. Change detection is one psychological process that would enable operators to possibly identify entities of concern when monitoring a particular region. Recall, in these experiments the entities could belong to one of six categories: friendly, assumed-friendly, neutral, assumed-hostile, hostile and unknown. The focus of the experiments is to compare different visual representations of these entity types and evaluate how they affect performance. In the control conditions, there were no separate representations for uncertain entities (i.e., there were no assumed entities). We compared this with two different ways of representing uncertainty: one where the assumed entities were different colours, the other where they were the same colour but hollow. We evaluated performance by examining participants’ accuracy in identifying the entity that changed.

Although these experiments are still currently being run, we can draw some conclusions as to what the pattern of results mean. First, we’ve found that, across at least two experiments, performance in the control and colour conditions was very similar. This suggests that adding a representation of uncertainty does not necessarily improve performance. However, we also found that performance *is* improved when using the hollow representation compared to the other two. This suggests that both entity representation and uncertainty have important roles to play in improving change detection performance.

The remaining two change detection experiments we are running aims to clarify whether and how the perceptual properties of the entities interact with uncertainty. However, the nature of the experiments limits the conclusions that we can draw about participants’ understanding of uncertainty. Indeed, these experiments seem to suggest that participants appear to be far less sensitive to uncertainty than we might expect or would predict from the literature. Therefore, the next series of experiments aims to *directly* explore how people represent, communicate and make decisions about uncertainty.

Below we summarise our planned approach below. The literature review (and common-sense) suggest that communication is the lynchpin in successful decision-making in a hierarchical organisation like the Navy. Thus, this line of work looks at communication in three ways that link together to mimic how decision-making happens onboard. Communication of uncertainty requires 1) the ``speaker’’ to successfully translate some internal representation of uncertainty to a message of some sort (words, numbers, pictures), 2) the ``listener’’ to translate that message back to an internal representation of uncertainty. The degree to which the communication is successful depends on how similar the representations between the speaker and listener are. The first two experiments explore 1) and 2) separately. Finally, depending a little on the results of these experiments, we plan to use computational modelling and more experimental work to explore approaches to combine uncertainty information from multiple people making separate judgements.

Table 1.

*Types of uncertainty examined in the proposed work.*

|  |  |
| --- | --- |
| Experiment 1 | Subjective uncertainty, Missing data, Predictive uncertainty |
| Experiment 2 | Subjective uncertainty, Predictive uncertainty, Uncertainty associated with measurement. |
| Experiment 3 | Inter-rater uncertainty, Uncertainty associated with measurement. |

**Speaker to Message**

In the literature review, we noted that most theoretical approaches in the categorisation and decision-making literatures argue that participants make decisions based on an integrated combinations of all the available information about a stimulus (Pothos & Wills, 2011). However, experimental work has shown that participants have a tendency (or bias) to use only a subset of the available dimensions (Noguchi & Stewart, 2018; Wills, Inkster, & Milton, 2015). Evidence that appears to indicate information integration may rather be an artefact of participants relying on different subsets of the available information. This observation raised a question: are participants (or indeed Naval operators) aware of which pieces of information are most important for classifying stimuli (entities) as one of several types (hostile, friendly etc.)? In other words, are people sensitive to predictive uncertainty? Predictive uncertainty can be thought of as the likelihood that an outcome follows a cue. For example, the likelihood that after eating rotting food (cue) you would get sick (outcome). Thus, in this experiment, we explore how sensitive participants are predictive uncertainty by exploring what they do in response to missing data.

Further, we look to see whether the stimulus representation matters. Often, representations of entities in military circumstances often combine all the available information into a single glyph. By automatically integrating the information into a single holistic representation might support participants in combining information that they would otherwise find difficult and thus, attend more to less predictive dimensions.

**The experiment in more detail**

Broadly, in this experiment, participants have to map entities (described by multiple pieces of AIS-like data) to either ``friendly’’ or ``hostile.’’ This experiment will have three stages: learning, test and report.

In the *learning* phase, on each trial participants will be shown the information concerning a particular entity. From that they will make a judgement as to whether it is friendly or hostile and then be given corrective feedback. We will provide five sources of information for each entity: the first source will be 90% predictive, the second 80%, the third 70%, the fourth 60% and the fifth

50%. These “information sources” will correspond to typical AIS data that might be used to make this judgements for actual vessels. For example, the most predictive dimension might be tonnage with 90% of large vessels being hostile. We plan to manipulate stimulus appearance between participants. In the separable representation, each dimension is displayed separately (perhaps as in Figure 1). This is similar to other ways that AIS data are commonly presented (see <https://www.marinetraffic.com>). In the integral representation, each dimension is displayed as part of a single glyph (as is common in the NATO symbology).

Table 2

*Stimuli assigned to Category A. Category B would be the exact opposite (i.e. exchange 1’s for 2’s and vice versa).*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Category | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 5 |
| A | 1 | 1 | 1 | 1 | 1 |
| A | 1 | 1 | 1 | 1 | 2 |
| A | 1 | 1 | 1 | 2 | 1 |
| A | 1 | 1 | 1 | 2 | 2 |
| A | 1 | 1 | 2 | 1 | 1 |
| A | 1 | 1 | 2 | 1 | 2 |
| A | 1 | 1 | 2 | 2 | 1 |
| A | 1 | 2 | 1 | 1 | 2 |
| A | 1 | 2 | 1 | 2 | 1 |
| A | 2 | 1 | 1 | 1 | 2 |

By the end of this phase, we expect participants to have inferred a strategy that allows them to perform reasonably well at this categorisation task. We expect to see that the majority of participants use only the most predictive dimension and ignore the rest.

In the second phase, we will present participants with the same types of stimuli. However, some of the dimensions will not be available to them. By removing dimensions (especially the more predictive dimensions) we can see whether participants gained any knowledge about the other dimensions. If a participant is sensitive to the predictive uncertainty of all the dimensions, they will select the second most predictive dimension, if they are not, they will likely select another dimension at random. Participants in this phase will have to again judge whether a particular vessel is hostile or friendly, and in addition they will have report their confidence in their judgements. We plan to manipulate how participants will give these confidence judgements: either as verbal descriptions ("Certain", "Likely", "Not likely", "Definitely not"), numerical judgements (the probability that the ship was whatever they said), or graphical (likely a pie chart).

Finally, participants will be asked to describe the strategy they used and why they chose that strategy in an open text box. This acts as a manipulation check for the modelling work that is required to determine participants’ strategies (Edmunds et al., 2018).

**Research questions**

**Q1:** By identifying the strategies participants use, we can determine whether participants integrate information from multiple sources to inform their decisions, or whether they base them on a subset of information.

**Q2:** By varying the presentation of information, we can see whether using an integrated representation of the stimuli improves the likelihood that participants integrate information or not. In other words, are the participants more likely to use different strategies when the representation of information is different?

**Q3:** By testing participants in the absence of some of the information, we can see whether they are sensitive overall to the predictiveness of the dimensions. For instance, if we remove the most predictive dimension (90%) then a sensitive participant should choose to categorise on the basis of the next predictive dimension (80%).

**Q4:** Finally, looking at the confidence judgements, we can see whether one method is better than the other in terms of accuracy or consistency. Thus, we can judge which communication format might be better for the “speaker” to document their subjective uncertainty.

A picture containing text, electronics, monitor, indoor

Description automatically generated

*Figure 1*. Example display of AIS data.

**Message to Speaker**

In the second phase, the focus is more on what people will do when faced with uncertain information and the best way of *receiving* uncertainty information. In other words, as stated in the statement of work, this experiments consider the different “types of decision” typified by the differing responses to uncertainty that a decision maker exhibits.

**Experiment in more detail**

In these experiments, we consider two types of uncertainty both manipulated within-subjects. We will again consider predictive uncertainty. On each trial, participants will see a single dimension that has one of five levels of predictiveness (e.g., 90% likely that large ships are hostile). Additionally, this information will be associated with one of five possible levels of informational uncertainty. Informational uncertainty is how likely to be accurate the information provided is (i.e., the likelihood that the ship is *actually* large). Informational uncertainty will be presented in one of three different ways (manipulated between-subjects) to match the confidence judgements made in the previous experiments (i.e., verbal, numeric or graphical).

On the basis of this information, participants must then decide whether they do not want any additional information or whether they would like to see additional information to make a final decision, or rather send someone to improve the accuracy of the data. If participants are sensitive to predictive uncertainty, we would expect them to be less likely to ask for more information for more predictive cues. Whereas, if participants are sensitive to informational uncertainty, we would expect them to be less likely to refine the input, the higher the accuracy of the reported data.

**Research questions**

**Q1:** Here, we can see whether participants understand the implications for different types of uncertainty for conducting the appropriate actions.

**Q2:** Further, we can see whether this is affected by how participants receive the uncertainty information (verbal, numeric or graphical). This is especially interesting if the optimum way of communicating information from the “listener’s” point of view is different from the “speaker’s”, as discovered in the previous experiment.

**Combining Information Across Operators**

The last section of work involves combining computational modelling and experimental work. The Navy consists of a hierarchical structure where incoming information is examined by many operators and must then be combined into a single picture. If Experiment 1 finds that people are poor at combining information, this would generalise to a superior office trying to combine information from many operators. As mentioned in the statement of work, a key challenge is to find novel and justified methods through which uncertainty may be represented (and communicated) to Defence decision-makers that intuitively supports their understanding and the aggregation of uncertainties to optimise decision-making.

Thus, in this section, we first determine the optimal performance that could be obtained via various tools or strategies. For instance, if there are 5 operators and they disagree on whether a vessel is friendly or hostile: how should we best combine this information so that the combination is correct most often. Then, in later experimental work we will look to see whether participants actually use these strategies and whether tools such as tallies or weighting algorithms can improve their performance.

**Research questions**

**Q1:** Here, we use modelling to find out how we should optimally combine information between multiple participants.

**Q2:** Additionally, we look to see whether participants actually come up with any of these solutions themselves.

**Q3:** Finally, we look to see whether adapting the appearance of the task can assist people in making more optimum choices with uncertain information.

**Some preliminary questions for DSTL**

* What (AIS?) information do you use to make these judgements?
* How is that information presented?
* How is information communicated between layers of the hierarchy?

References

Edmunds, C. E. R., Milton, F. & Wills, A. J. (2018). Due process in dual process: Model-recovery simulations of decision-bound strategy analysis in category learning. *Cognitive Science, 42,* 833-860.

Noguchi, T., & Stewart, N. (2018). Multialternative decision by sampling: A model of decision making constrained by process data. *Psychological Review*, *125*(4), 512–544. doi: 10.1037/rev0000102

Pothos, E. M., & Wills, A. J. (Eds.). (2011). *Formal approaches in categorization*. Cambridge University Press. doi: 10.1017/cbo9780511921322

Wills, A. J., Inkster, A. B., & Milton, F. (2015). Combination or differentiation? two theories of processing order in classification. *Cognitive Science*, *80*, 1-33. doi: 10.1016/j.cogpsych.2015.04.002