



UNDER-TRUST IN DECISION SUPPORT SYSTEMS

Introduction

It is often assumed that two heads are better than one, but reliance on decision aids is often inappropriate. Decisions to rely on an aid are thought to be based on a comparison between the perceived reliability of own performance and that of the decision aid. Unfortunately, perceived reliabilities are unlikely to be perfectly calibrated. This can result in inappropriate decisions to rely on advice.

Hypotheses

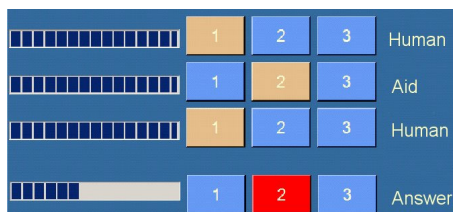
Based on the literature we expect under-estimation of own reliability and that of the aid. But the question remains whether there are differences between individuals and between self the aid. We test three hypotheses:

H1: Under-estimation of the decision aid is more prevalent than of the self.

H2: Under-estimation of the self decreases after practice; that of the decision aid persists.

H3: Unreliability of the decision aid is attributed differently compared to own unreliability.

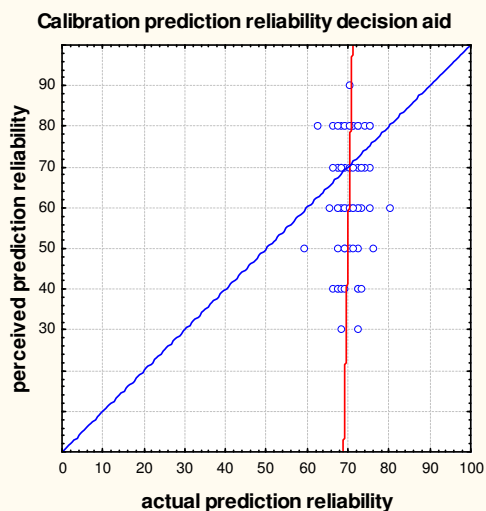
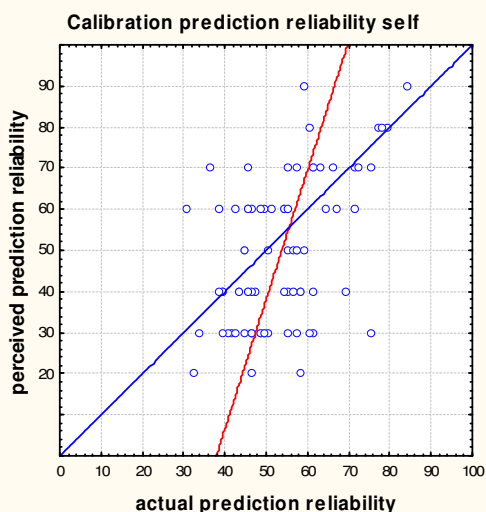
Method



The experimental environment is shown above. It consists of a prediction, advising, selection, and feedback phase, respectively. In a laboratory experiment with 40 participants, we studied whether calibration improves after practice, whether calibration of own reliability differs from calibration of the aid's reliability and whether unreliability of the aid is attributed differently. We did this by measuring performance as well as asking questions after trials.

Results

The results of the measured calibration in the prediction and advising phase are shown below. The results of the selection phase are omitted for reasons of brevity.



Calibration learning effects were measured within subjects by means of a comparison between two trials. For the first trial participants under-estimated own performance with 5%, $t(39) = -2.16$, $p < .05$. Under-estimation disappeared in the second trial, $t(39) = -1.46$, $p > .05$. Participants under-estimated the performance of the decision aid in the first trial with 7%, $t(39) = -2.47$, $p < .05$, as well as in the second trial (8%), $t(39) = -3.79$, $p < .01$.

On a linear scale [-3, 3], compared to prediction of own reliability ($M = .41$)

the prediction reliability of the decision aid is least attributed to temporary factors ($M = .05$), $t(79) = 2.02$, $p < .05$. Compared to human prediction reliability ($M = -0.79$) the prediction reliability of the decision aid is least attributed to external factors ($M = -1.09$), $t(77) = 1.66$, $p > .05$.

Conclusion

Our results suggest that calibration is imperfect and overall people think reliability is worse than it actually is.

Overall negative feedback seems to receive more weight than positive feedback. Although under-trust was more prevalent (H1), we found both pessimists that seem to over-weigh negative feedback as well as optimists that seem to under-weigh negative feedback.

The direction of error in calibration does not differ between self and aid, but our results suggest that the magnitude of under-estimation differs, especially after practice.

Calibration of the own prediction reliability improves after practice and under-trust disappears. Calibration of the reliability of the decision aid, however, does not improve after practice and under-trust remains (H2). Also we found that under-trust in reliance decisions does not disappear.

Unreliability of the decision aid is less likely to be attributed to temporary external factors (H3). The asymmetry in attribution and calibration may explain under-reliance on decision aids that is often found.

Acknowledgments

This research is funded by the Royal Dutch Navy (pr. nr. V206).

TNO Defence, Security, and Safety
P.O. Box 23
3769 ZG Soesterberg
The Netherlands